I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11958

The Elasticity of Taxable Income: A Meta-Regression Analysis

Carina Neisser

NOVEMBER 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11958

The Elasticity of Taxable Income: A Meta-Regression Analysis

Carina Neisser
ZEW Mannheim, University of Mannheim and IZA

NOVEMBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Elasticity of Taxable Income: A Meta-Regression Analysis*

The elasticities of taxable and broad income are key parameters in tax policy analysis. To examine the large variation in estimates found in the literature, I conduct a comprehensive meta-regression analysis using information from 51 studies containing 1,448 estimates. Heterogeneity in reported estimates is driven by regression techniques, sample restrictions and variations across countries and time. Moreover, I provide descriptive evidence of the correlation between contextual factors and the magnitude of an elasticity estimate. Selective reporting bias is prevalent in the literature and the direction of reporting bias depends on whether or not deductions are included in the tax base.

JEL Classification: C81, H24, H26

Keywords: elasticity of taxable income, income tax, behavioural response, meta-regression, analysis

Corresponding author:
Carina Neisser
ZEW Mannheim
L 7, 1
D-68161 Mannheim
Germany
E-mail: carina.neisser@zew.de

* I gratefully acknowledge financial support by MannheimTaxation (MaTax) Science Campus. I would like to thank David Agrawal, Philipp Dörrenberg, Paul Hufe, David Jaeger, Andreas Lichter, Max Löffler, Claus Kreiner, Andreas Peichl, Jörg Paetzold, Emmanuel Saez, Sebastian Siegloch, Holger Stichnoth, Christian Traxler, Johannes Voget, Mazhar Waseem, Mustafa Yeter and participants at the LAGV 2016, IIPF 2016, PF Seminar (U Mannheim), ZEW Public Finance 2017 and IEB Taxation Workshop 2017 for their helpful comments. Philipp Kollenda provided excellent research assistance.

The elasticities of taxable (ETI) and broad income (EBI) are key parameters in tax policy analysis. The amount of literature in this area has grown substantially over the last two decades. Despite the importance and the large body of literature estimating this parameter, there is little consensus on the magnitude of the elasticity that should be used in economic policy analysis and there are various explanations of why these estimates differ.[1] The majority of estimates lies between 0 and 1 with a peak around 0.3 and an excess mass between 0.7 and 1. Given that we know that behavioural responses to taxation are not structural parameters, the goal of this paper is to identify and assess different explanations for the pattern of estimates found in the empirical literature by applying meta-regression techniques.

Taxable and gross elasticities summarize different types of behavioural responses to income taxation such as real responses (e.g. labour supply adjustments), tax avoidance (e.g. claiming deductions or (legal) income shifting between tax bases) and illegal tax evasion behaviour. Elasticity estimates are based on to the net-of-tax rate (NTR), which is the after-tax share that an individual receives from reporting another unit of income. The magnitude of behavioural responses to tax rate changes is of major importance for the design of tax and transfer policy. The ETI serves as a behavioural parameter in optimal taxation models (e.g. Mirrlees, 1971; Diamond, 1998; Saez, 2001; Piketty and Saez, 2013) and under certain assumptions, it is also a sufficient statistic for dead-weight loss calculation (see Feldstein, 1999 or Chetty, 2009). Since Feldstein (1995), a large body of empirical work estimating taxable income responses has emerged. Much of this work deals with the US. In recent years non-US studies based on different identification strategies and datasets have also been published (compare Saez et al., 2012 for a survey).

There are many reasons why estimated elasticities can vary. Giertz (2007, 2010*b*) uses various time periods (involving different tax reforms) and different datasets to estimate elasticities for the US. He applies various estimation techniques and his results, reveal

[1] 'Elasticity of taxable income' is often used as an umbrella term that involves also the elasticity of gross income. I use the term elasticity of the income base as a synonym for all types of elasticities (e.g. adjusted gross and taxable income) and I differentiate between before (BD) and after deduction (AD) elasticities in the later analysis.

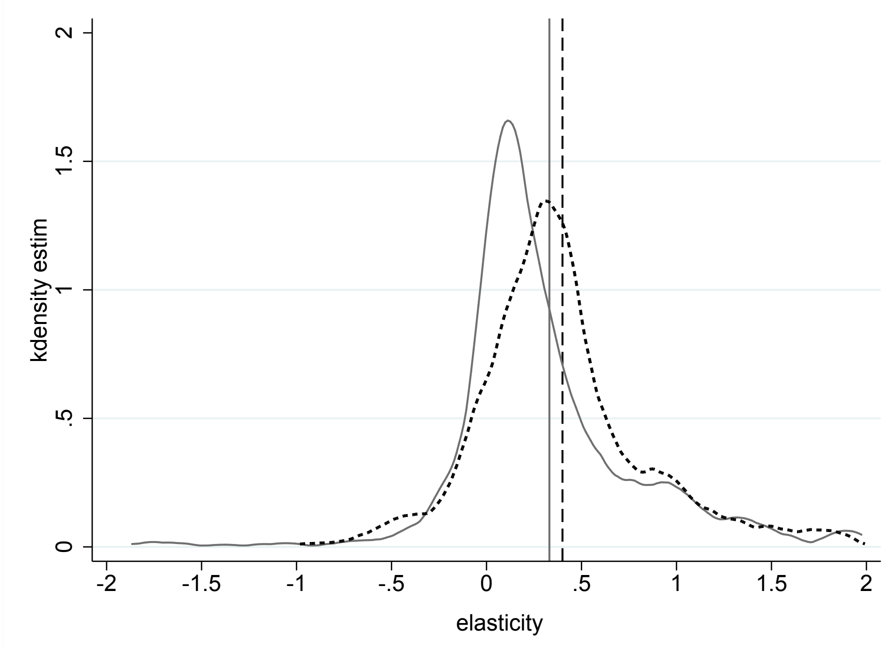
a considerable heterogeneity in the size of elasticity estimates. A precise estimation of changes in (taxable) income resulting from marginal tax rate changes is challenging because of confounding factors such as mean reversion, heterogeneous income trends and, most importantly, because in a progressive tax system income and marginal tax rates are jointly determined. Various strategies have been applied to overcome these problems. Auten and Carroll (1999) and Gruber and Saez (2002) use an instrumental variable (IV) approach along with income control variables. Recently, more developed estimation methods involving different instruments and control variables have been applied (e.g. Blomquist and Selin, 2010; Weber, 2014*b* or Burns and Ziliak, 2017).[2]

Another branch of research tries to explain the behaviour of taxpayers and provide reasons why elasticities are not immutable parameters. Slemrod and Kopczuk (2002) and Kopczuk (2005) highlight the fact that the ETI is considerably larger in tax systems with more deduction possibilities and can therefore be controlled by policy makers. Much of the evidence is based on self-employed and/or high-income taxpayers given their larger range of opportunities to adjust their (taxable or gross) income (e.g. Kreiner et al., 2014; Kreiner et al., 2016; Le Maire and Schjerning, 2013 or Harju and Matikka, 2016). Kleven et al. (2011) and Kleven et al. (2016) stress that third party information reporting influences the magnitude of behavioural responses.

In order to reconcile the variation in the estimates, I provide a comprehensive quantitative survey of the empirical literature. I collect 1,448 estimates extracted from 51 studies. To disentangle real and reporting responses by individuals, I explicitly differentiate between behavioural responses that are based on income concepts that consider or do not consider deductions. I allocate all reported elasticities to two subsamples: before (BD) and after deduction (AD) elasticities. Figure 1 plots the distribution of elasticities of the BD and AD subsamples. The vast majority of estimates (90%) lies within an interval of -1 and 1 with a strong propensity to report estimates between 0 and 1. There is a peak around 0.3 and an

[2] I only consider Difference-in-Differences and Instrumental variable estimations and do not cover bunching or share analysis because resulting estimates are not comparable to each other. Moreover, I only focus on income taxation.

Figure 1: Distribution of elasticities



Note: The distribution of before deduction (BD) elasticities are displayed as a solid line and the corresponding vertical line highlights the mean of 0.300. The distribution of after deduction (AD) elasticities are displayed with a dashed line and the corresponding mean of 0.389 is highlighted with the vertical dashed line. Both figures are based on an an epanechnikov kernel with a bandwidth of 0.072.

excess mass between 0.7 and 1.

To identify and quantify sources of heterogeneity, I apply a meta-regression analysis. Different types of sources (e.g. empirical strategy or country) are recorded and differences between elasticity estimates are quantitatively examined across the entire empirical literature. Importantly, my meta-analysis provides a replicable statistical framework for summarizing and assessing the full range of empirical evidence. Besides a separate analysis for BD and AD elasticities, I check for the influence of the underlying estimation technique (regression techniques, income control, difference length and weighting by income) on primary elasticities. Moreover, I investigate whether sample restrictions affect the estimates. Variations across countries and time are analysed. To provide new evidence, I add contextual variables (country year) to my analysis. More precisely, I show correlations between tax- and economy-related characteristics (tax reform related characteristics, income inequality, business cycle effects and third party information reporting) and estimated elasticities. Apart from this, I explicitly test whether selective reporting bias is prevalent in the literature.

My meta-regression analysis generalizes findings across the entire empirical literature and corroborates the following previous results. First, elasticities that include deductions are more elastic compared to elasticities that are based on gross income. Second, elasticities are sensitive with respect to the underlying empirical strategy. Results already obtained by Giertz (2010*b*), who only studies the US, are confirmed. Third, sample restrictions mostly affect AD results. Finally, estimation and publication decade as well as the country under observation have an impact on estimated elasticities.

My study shows that mainly before deduction elasticities that are correlated with contextual variables. A notable exception is the correlation between AD elasticities and top income shares. This analysis highlights the fact that estimated elasticities are affected by past as well as current (tax) policy and that the underlying context matters when interpreting these elasticities.

Finally, this paper is related to the literature on the so-called 'file drawer problem' (Rosenthal, 1979). I show graphically and statistically that the literature suffers from selective reporting bias. There is a tendency to report significant results more often. The existence of 'p-hacking' is more prevalent among published articles compared to working papers. This pattern is more pronounced among AD elasticities. In addition, results that are in line with theory are reported more frequently. There is an upward reporting bias for BD elasticities. For AD elasticities, the reporting bias goes in both directions, while the downward bias appears to be more dominant. There is an aversion to reporting negative estimates and estimates above 0.4 and in particular above 1. In general, AD elasticities are more likely to get reported if they lie in a range between 0 and 0.4.

My findings have important policy implications. Behavioural elasticities towards taxation are key parameters for determining optimal marginal tax rates and the welfare costs of taxation. Biased elasticities result in skewed policy implications. Under certain assumptions, I demonstrate that an aversion of an AD elasticity above 0.4 leads to an aversion against optimal top tax rates below 0.52. I highlight the role of administrative data and suggest improvements that might increase transparency and reduce selective reporting bias.

This paper contributes to the literature by giving an objective overview and rigorous analysis of the empirical evidence on behavioural elasticities with respect to income taxation. I examine the systematic impact of various factors on the reported elasticity estimates. Although the ETI literature has been reviewed by Saez et al. (2012), I am not aware of any meta-regression analysis of taxable income elasticity estimates. My study follows a strand of literature that applies meta-regression analysis (see Christensen and Miguel, 2016 for a review).[3] Moreover, this is the first study that relates existing empirical evidence to contextual factors.

The remainder of this paper is structured as follows. In Section 1, I explain the meta regression model and I describe the data collection process (1.2). In Section 2, I outline a basic framework to discuss empirical challenges in the literature on taxable income elasticities (2.1) and provide explanations of defined sources of heterogeneity (2.2) along with descriptive statistics (2.3). In Section 3, I provide and discuss the baseline results and descriptive evidence on the influence of contextual factors on elasticities (3.2). In Section 4, I examine selective reporting bias. I provide evidence (4.1) and discuss these findings (4.2). Section 5 concludes.

1 Meta-regression Framework and Data Collection

I follow standard meta-regression analysis techniques (e.g. Card et al., 2010, 2018) and present the meta-regression framework and the estimation technique employed in section 1.1.[4] In section 1.2, I describe the data collection, the applied exclusion restrictions and the final dataset.

[3] Card and Krueger (1995) and Card et al. (2010, 2018) are three examples that look into the field of labour economics. Havránek (2015) examines the literature on intertemporal substitution elasticities and Lichter et al. (2015) study labour demand elasticities. Moreover, there is a large body of literature on publication bias (see Rothstein et al. (2006) for a review and Brodeur et al. (2016) for an application).

[4] See Feld and Heckemeyer (2011) for a methodological review and Lichter et al. (2015) for a different but similar application.

1.1 Meta-regression Model

The meta regression model is given by

$$\zeta_{is} = \zeta_0 + \beta X_i + \delta Z_{is} + \epsilon_{is}, \tag{1}$$

where ζ_{is} represents the i-th estimate of the respective elasticity collected from study s. ζ_0 denotes the intercept, X_i and Z_{is} represent study and estimate-specific variables respectively, and ϵ_{is} is the sampling error. Since the variances of collected estimates are heteroscedastic, it is preferable to estimate the model using Weighted Least Squares (WLS) rather than through an OLS estimation. I use the inverse of the error term variance of an individual estimate $V(\hat{\zeta_{is}}) = \sigma_{is}^2$ as analytic weights. Hence, I give observations with smaller variances a larger weight and greater influence on the estimates since precision can be seen as an indicator of quality. Standard errors are clustered at the study level to control for study dependence in the estimates.[5]

1.2 Data Collection

A comprehensive review and examination of the ETI literature provided the data for the meta-analysis.[6] As a first step, I searched Google Scholar and IDEAS RePEc using the following search terms: 'elasticity of taxable income', 'eti', 'taxable income', 'new tax responsiveness' and 'tax elasticity'. In addition, I relied on a survey by Saez et al. (2012) to identify relevant studies published prior to 2011 and I cross-checked these with the reference list of all previously identified papers. The search process lasted from February 2015 to

[5]To test the robustness of the results with respect to the underlying weights, I conduct various regressions (see (online) Appendix F): (1) a simple OLS, (2) Random effects meta-regression technique, (3) a WLS with weights that are based on the inverse of the share of observations per study in relation to the full sample and (4) a WLS with weights that account for the sample size of each study. To check whether clustering in the meta-analysis produces misleading inferences, I apply a wild-cluster bootstrap procedure proposed by Cameron et al. (2008) for improved inference with only a few clusters.

[6]The meta analysis follows reporting guidelines proposed by Stanley et al. (2013). A list of people who have coded and checked the data, a list of identified but not-included studies and estimates or a list of all included estimates plus sources is provided upon request. I checked only English and German articles. In June 2016, all working papers were re-checked for any updates. I ignore the literature on responses to corporate taxation or capital gains and I rely on estimates that are based on commonly used income concepts and income tax changes as a source of variation.

December 2015 and I identified 203 potential studies.

In the second step, I applied certain exclusion criteria to determine the final sample of studies. I only considered studies that measure responses to income taxation and exploit differential changes in tax treatment following tax reforms that are based on Differences-in-Differences (DID) and Instrumental Variables (IV) estimations. I did not cover share/time-series analysis and bunching because the resulting estimates are not comparable to each other. I only coded studies that provide their own empirical estimates and rely on commonly used income concepts as described below. Based on this sample, I found 37 studies that were published in a peer reviewed journal. Additional working papers increased the number of articles to 51.[7]

In the third step, I collected every estimate derived from a different specification (so-called multiple sampling) so that they are different with regard to the defined sources of heterogeneity (e.g. income concept or sample restrictions). I collected all point estimates, corresponding standard errors, number of observations and type of control for heteroscedasticity and autocorrelation. Additional information on journal, year of publication, source (where to find a particular estimate), country and time period is coded. In a fourth step, I restricted the final dataset. I considered only estimates that provide a standard error or t-statistic. This reduced my number of observations from 1,514 to 1,448.

Finally, I collected all necessary study characteristics, which I will explain in the next section. Additional information on contextual factors such as tax reform and economic characteristics are collected and merged with the dataset (see Table 1 for an overview).

[7] In the (online) appendix A, I provide a table with studies included in the sample and I show the distribution of estimates by study. On the one hand, adding unpublished papers to the meta-sample might lower the quality of included estimates but, on the other hand, most working papers are more recent and use better datasets and improved estimation techniques. It should be noted that this meta study is only as good as the studies on which it is based and there might be variation among the studies that cannot be reflected by the coded variables.

2 Elasticity of Taxable Income

I briefly explain the concept of taxable income elasticities and state the empirical challenges. For a detailed discussion, please refer to an excellent survey by Saez et al. (2012). I outline various reasons why elasticity estimates differ and describe the coded characteristics. The dependent variables are summarized such that they belong either to the before or after deductions subsample. Finally, I provide some descriptive statistics.

2.1 Empirical Challenges

The (taxable) income literature uses an extension of the traditional labour supply model. Individuals maximize a utility function $u(c, z)$, where z is income and c consumption. An elasticity of the income tax base measures the responsiveness of income to changes in the net-of-tax rate (NTR) - defined as one minus the marginal tax rate. This is the percentage change in income in response to a one percent increase in the NTR. An increase in the marginal tax rate reduces the NTR, which in turn reduces taxable income. Hence, the expected elasticity should be positive.[8]

Several conditions must hold in order to estimate behavioural responses correctly. First, only the marginal tax rate changes while the tax base changes are kept constant. In reality, the underlying tax base changes as often as the tax rate itself. To rule out any tax legislation-induced tax base effects, the broadest definition and, therefore, an 'artificial' tax base across years is used.

Second, an ideal empirical setting would compare two randomly selected but similar groups before and after the introduction of a policy change where one group experiences a tax rate change (=treatment) while the other does not (=control). The parallel trend assumption needs to be fulfilled in order to obtain an unbiased estimate. Typically, treatment and control groups are defined based on their income. This is problematic since non-tax related factors as well as tax changes are correlated with income.

[8] Information about estimated income effects is rarely available (e.g. Gruber and Saez, 2002 or Bakos et al., 2010) so, I ignore them and assume that compensated and uncompensated elasticities are equal.

Third, in a progressive tax system the marginal tax rate τ and income z are jointly determined and tax rates increase automatically if an individual faces a (non-tax related) positive income shock and potential income responses are (wrongly) captured by the ETI. To establish a clear link between income and tax rate changes, researchers mostly use a two-staged least squares estimator as a regression technique and instrument for the change in NTR.

Fourth, different income growth rates across the population (e.g. larger income growth for high-income earners) and reversion to the mean further aggravate a 'clean' estimation. These shocks influence the shape of an income distribution and they need to be incorporated in an empirical framework. Secular and heterogeneous income trends (e.g. larger income growth at the top), lead to an upward bias and mean reversion might go into both directions depending on the type of income shock. Given the presence of such non-tax related factors, since Auten and Carroll (1999) researchers include income control variables. The most standard regression specification is derived as:

$$log\left(\frac{z_{it}}{z_{it-k}}\right) = \zeta log\left(\frac{1-\tau_{it}}{1-\tau_{it-k}}\right) + \delta f(z_{it-k}) + \theta X_{it-k} + \mu_t + \epsilon_{it}, \tag{2}$$

where k is the chosen difference length and $t-k$ denotes the base-year. X_{it-k} is a vector of control variables. Time dummies μ_t control for any omitted variables in differences that are the same on average for all individuals. $f(z_{it-k})$ denotes the income control in order to capture non-tax related income trends.

Fifth, tax reforms provide variation in marginal tax rates used for identification and changing definitions of the tax base are ignored. A tax change is assumed to be exogenous if it is not systematically correlated with other developments that affect economic measures like GDP. For instance, a tax policy that reduces taxes because policy makers are anticipating a recession is clearly endogenous (Romer and Romer, 2010). Even previous tax policy can affect current elasticities. Finally, many tax reforms do not target a single income group and tax rates change by different magnitudes throughout the income distribution.

2.2 Sources of Heterogeneity in Estimated Elasticities

Many factors influence the size of an estimate. To assess the relevance of different explanations, I define various dimensions of heterogeneity: (1) income concept, (2) estimation techniques, (3) sample restrictions, (4) publication characteristics, variation across countries and time and (5) contextual factors. Dimension (1) to (4) are collected from primary studies while dimension (5) is based on external data sources. There are more dimensions of heterogeneity worth investigating, such as the role of income effects, restrictions on demographics (e.g. gender) or tax-related characteristics (e.g. no alternative minimum tax (AMT)) and even certain control variables like education. However, a limited number of estimates account for these things which makes it not possible to test for them. Table 1 provides an overview of all included characteristics and I describe each coded variable in greater detail in the (online) appendix C.2.[9]

Income Concept. An elasticity of the income tax base measures the responsiveness of income to marginal tax rate changes. Since an elasticity is a function of the definition of the tax base, a central question is what type of income should be used as a dependent variable? Ideally, I would like to observe a comparable and uniformly defined income across all studies. This is impossible even for conceptually equal income concepts like taxable income. The exact definition varies from country to country and even within a country over time. In many studies, details of the tax simulation model are missing and lack of transparency on how income variables are constructed makes it difficult to compare estimates with each other.

Researchers mainly use taxable income, adjusted gross income and total income. Total income (= gross or broad income) is the sum of all income. Subtracting specific deductions (e.g. retirement plan contributions), yields adjusted gross income. Taxable income is calculated as adjusted gross income minus personal exemptions and itemized deductions.[10]

[9] I provide additional descriptives in the (online) appendix C and more details on the distribution of elasticities and details on explanatory variables in appendix D.

[10] Although capital gains are part of taxable income, only a few studies explicitly mention that they subtract capital gains. In addition, most researchers remain salient on whether or not they apply a constant tax base

Scandinavian studies look at earned income since these countries apply a dual income tax system which taxes labour and capital at different rates.[11]

Behavioural responses towards taxation can take many forms including changes in labour supply (participation and working hours), tax avoidance (changing the timing of income/ transactions, changes in the extent of spending on tax deductible activities, e.g. donations, or even claiming questionable deductions) and tax evasion (understating income, claiming unjustified deductions). The distinction between whether or not an income concept considers deductions is crucial, since it determines the range of responses. Real responses can be captured with a before-deduction elasticity while an after-deduction elasticity captures a broader range of responses including avoidance behaviour. Tax evasion affects both types of elasticities. Hence, I only distinguish whether or not the dependent variable in primary studies/ estimates consider deductions and I allocate all reported income concepts to two subsamples: before (BD) and after deductions (AD).

It is important to keep in mind that real responses such as labour supply responses but also avoidance and evasion behaviour depend on an individual's preferences for work and leisure and they can to some extent be influenced by the tax system in place (Slemrod and Kopczuk, 2002). Kopczuk (2005) shows how the ETI varies with its tax base. While the ETI (=AD elasticity) is considerably larger in a tax system with more deduction possibilities, it can also be lower in a country with a high degree of third party information reporting (e.g. exchange of information between employer and tax authority) (Kleven and Schultz, 2014). Hence, the magnitude of the ETI is influenced by the design of the tax system itself and is therefore a policy choice (Slemrod, 1995).

Estimation techniques. I define three distinctive features with respect to estimation techniques that influence the ETI: (a) regression technique, (b) income control, (c) difference

approach. Both things influence the definition of taxable income and therefore the results.

[11] In the (online) appendix, Table 8 shows the distribution of elasticities by reported *income concept* within the dataset. Additional descriptives are provided. As a sensitivity check, I run the estimations on a subsample of the dataset and look only at taxable income elasticities (see Table 19). These results remain unchanged compared to estimation results that consider all AD estimates.

length and (d) weighting by income.

I categorize five *regression techniques*. Since income and marginal tax rates are jointly determined, almost all approaches follow an Instrumental Variable (IV) procedure. They essentially differ in the way they instrument for the NTR. The most standard approach is defined as 'IV: mechanical tax rate changes'. The idea is that this change in net of tax rates is free of any behavioural responses and represents only mechanical changes that can be used as an instrument for the NTR. It was first implemented by Auten and Carroll (1999) and Gruber and Saez (2002). To construct mechanical tax rate changes, one uses income from base year $t-k$ and assumes that it remains the same in year t. Applying tax rules for year t yields a mechanical (sometimes called predicted or synthetic) tax rate.

Finding instruments that satisfy all relevant conditions to receive consistent estimates is a major problem. More developed estimation methods involving different instruments and control variables have been applied. The second estimation technique is called 'IV: (lagged) mechanical tax rate changes'. Weber (2014*b*) argues that mechanical tax rate changes mentioned above should be lagged in order to fulfil the exclusion restriction. Her approach makes it possible to deal with serially correlated transitory income shocks. Besides Weber (2014*b*), different instruments have recently been developed such as in Blomquist and Selin (2010), Burns and Ziliak (2017), Gelber (2014) or Matikka (2016). I summarize all other types of instruments in a third category (IV: other).

The earliest method, namely a basic Difference-in-Differences (DID) approach, uses a defined treatment and control group without any instruments and income controls. As explained before, it is hard to define a clean treatment and control group and to disentangle income growth driven by tax and non-tax effects, particularly if the treatment status is based on income. In the case of tax cuts for upper-income groups, secular changes in income (e.g. larger income growth at the top), lead to an upward bias and mean reversion might go in both directions depending on the type of income shock. Difference-in-Differences (DID) with a dummy variable as an instrument represents another category. This is a conventional DID approach in which the NTR is instrumented by the interaction of the after-reform and treatment group dummy. This is similar to Feldstein's (1995) tabulated DID approach, but estimated in a regression framework that allows for additional control variables (Moffitt and Wilhelm, 2000).

Income control variables are additional explanatory variables to capture non-tax related income growth. Administrative tax datasets offer precise information about a taxpayer's income and deductions. However, socio-demographic information is limited. This further limits the estimation possibilities such that researchers rely on income controls to approximate a taxpayer's wealth. The success of income controls depends on the extent of year-to-year mean reversion and the stability of the underlying income distribution. It is therefore unclear what type of income control performs the best.

I define five generations of income controls. First, there is the use of no additional income control variables (none). Studies published prior to 2000 use no income controls and most studies estimate a specification with no income controls as a sensitivity check. The second generation covers studies that use only the log of base-year income control $ln(z_{i,t-k})$ (Auten and Carroll, 1999). Following Gruber and Saez (2002) researchers use more sophisticated income controls like a spline of log base-year income. A spline allows us to control for non linear income trends across income groups by dividing income groups into deciles. Kopczuk (2005) argues that using only base-year income and some flexible function is not sufficient. He explicitly distinguishes between permanent and transitory income components and proposes two types of income control variables: the log of lag base-year income $ln(z_{i,t-k-1})$, that allows to control for an individual's rank in the income distribution and therefore for the permanent income level and transitory income trends are captured by using the deviation between log base-year and log lag base-year income $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$. The last generation covers every *other* (non-standard) income control used in the literature, e.g. grouping income control as used in Burns and Ziliak (2017) or the income spline defined as in Gelber (2014). A potential risk of using income controls might be that they absorb too much identifying variation.

All studies apply a 'First Difference' estimation strategy with a varying *difference length* to eliminate the impact of unobservable time-invariant characteristics. An estimate is either based on a 1-year, 2-year, 3-year or of 4 and more years specification. Most estimations use a 3-year time window such that researchers relate income and marginal tax rates e.g. from 2001 to 2004 and Kleven and Schultz (2014) show graphically that three-year lengths are sufficient to account for behavioural adjustments. One might think that the longer the time window the larger the behavioural response. However, the timing, announcement and implementation of underlying reform(s), individual speed of understanding as well as an individuals' ability to adjust their income has an effect on the size of behavioural adjustments. Since many tax reforms are phased-in over several years, an estimate is only a combination of short-, medium- and long-run responses (Weber, 2014*b*).

Since weighted elasticity parameters reflect the relative contribution to total revenues, regression results are sometimes *weighted by income* (Gruber and Saez, 2002.[12] If, as expected, responses do not vary by income, weighting the estimates by income will not affect elasticity estimates. However, it seems reasonable to assume that behavioural responses are not homogeneous across the income distribution. Weighted results account for the fact that high-income taxpayers tend to exhibit larger responses. Typically, these weights are censored at the top (e.g. at 1 million $) and are not free of criticism, because income itself is endogenous (Weber, 2014*b*). For instance, if high income taxpayers face an income shock, this will bias the results.

Sample restrictions. Researchers apply sample restrictions to avoid the problem of outlier and to conduct sensitivity analysis. Since mean reversion is pronounced at the bottom of the income distribution, the income distribution is restricted from below. The truncation of the income distribution affects the comparison group that is used to explain counter factual income trends. However, if this improves or impairs, the identification depends on non-tax related income growth across the income distribution. I coded whether *income cutoffs* are used, and if so, the corresponding threshold. These thresholds are re-calculated

[12] Similar to missing details regarding whether or not capital gains are included in the income concept, it often remains unclear by what type of income estimates are weighted.

in US-Dollar. To account for mean reversion at the beginning and end of an individual's working life, researchers apply an *age cutoff* to limit the sample to the working population and to exclude pensioners.[13]

Publication characteristics and variations across countries and time. To account for potential differences, I control for whether or not an estimate is reported in a peer reviewed journal or in a working paper. Given the research process, I include different categories for *publication decade* ((1) <= 2000, (2) 2001-2010; (3) >2010) as controls. Publication decade does not necessarily coincide with the timing of a tax reform. To identify a potential development over time that is not directly related to any type of methodological progress but rather related to tax policy at a given time, I include *estimation/ data decade* as a control. For a particular estimate, I calculate the mean of first and end year of the underlying data period ('mean year of observation') and assign the corresponding decade: 1980s, 1990s or 2000s. Countries are summarized in different *country groups* (1) USA, (2) Scandinavia (Denmark, Norway, Sweden), and (3) other countries (Canada, Finland, France, Germany, Hungary, Netherlands, New Zealand, Poland, Spain).

Contextual Variables. There is evidence that behavioural responses to income taxation are related to contextual factors. Fack and Landais (2016) show that the magnitude of behavioural responses is extremely sensitive to the level of tax enforcement. Giertz (2007) exploits different tax reforms and estimates heterogeneous elasticity estimates. Kleven and Schultz (2014) find that behavioural elasticities are larger when estimated from large tax reform episodes. Both studies highlight the fact that an elasticity is sensitive to the underlying tax reform used for identification. Similar to Chetty et al. (2011) and Chetty (2012), Kleven and Schultz (2014) also show that a more salient tax reform is more likely to overcome optimization frictions.

Tax reforms are necessary to generate variation that can be exploited. A reform does

[13]In the (online) appendix F, I provide estimation results that account sample restrictions with respect to *marital status* and *employment type*.

not happen in a single year, nor is it easy to tell exactly which income group is affected. Moreover, most estimates are based on a data period with more than one single change in tax law. This makes it difficult to account for tax reform characteristics in the meta analysis. Therefore, I only account for two factors. First, the *introduction of a top tax bracket*. Since such a reform is more salient and the affected tax group is the most responsive one, this might lead to higher estimates.

Economic characteristics shape behavioural responses to taxation as well. To account for income inequality within an economy, I include the *Gini coefficient* (disposable income, post taxes and transfers). It is based on the comparison of cumulative proportions of the population against cumulative proportions of income they receive. It ranges from 0 in the case of perfect equality to 100 in the case of perfect inequality. In addition, I include a measure of the share of pre-tax national income that is held by the *top 1%* and *top 10%* as contextual variables in my regression. An increase in inequality might be the result of past tax cuts for high-income tax payers. Hence, larger estimates might not be the result of larger responses, but rather of a widening in the income distribution that is ('wrongly') captured by estimated elasticities.

Hargaden (2015) provides evidence of a weaker behavioural response during a recession and therefore highlights the role of business cycle fluctuations. To account for a given economic situation, I add the respective *unemployment rate* as a contextual variable in my regression.

Third party information reporting plays a key role in tax compliance and a country's overall tax take. Kleven et al. (2011) find that the overall tax evasion rate is very small in Scandinavia because almost all income is subject to third party information reporting. It has been shown in the literature that tax enforcement is strong whenever third-party information reporting (e.g. through the exchange of information of employers or banks and tax authorities) is in place. I include two variables as a proxy to check for its influence. First, the *fraction of self-employed* workers within a country. Traditionally, self-employed taxpayers provide most of the necessary information to tax authorities themselves. I expect a positive relationship between elasticities and the share of self-employed workers within an economy.

As a second measure, I include the share of *modern taxes per GDP* to proxy for the share of tax revenue that are exposed to third-party information reporting compared to the overall tax take. Kleven et al. (2016) distinguish between what they call traditional and modern taxes. Unlike traditional taxes that rely on self-reported information, modern taxes rely on third-party information. Modern taxes are defined as personal and corporate income taxes, value-added taxes, payroll taxes and social security contributions whereas traditional taxes are all other taxes (e.g. inheritance tax). Modern taxes play a crucial role in the economic development of a country and there is a strong positive correlation between GDP per capita and modern taxes to GDP. I expect a negative correlation between reported elasticities and modern taxes to GDP ratio.

2.3 Descriptive Statistics.

Table 1 provides an overview of the collected information to explain differences in elasticity estimates. As already mentioned, I divide the meta-sample in two subsamples depending on whether the underlying income concept accounts for deductions. The before deductions subsample consists of 852 observations collected from 38 studies and the after deduction subsample of 596 observations from 37 studies.

Around 60% of the estimates refer to a regression technique that uses mechanical tax rate changes as an instrument. One third of estimates use the log of base year income (Auten and Carroll, 1999) as an income control. Most estimates either use a difference length of three years or consider a short time window of one year. One third of all primary estimates are weighted by income. Almost half of the estimates apply an age cutoff. The vast majority of estimates use an income cutoff.

More than 70% of the estimates were published in a peer-reviewed journal and most estimates use US data. The mean year in datasets used by primary studies is 1993 and one third of all estimates are estimated in the 1980s, 1990s and 2000s respectively. The meta sample covers studies released between 1987 and 2016. The mean year of publication is 2009.

One third of all collected estimates are based on a data length that lasts for less than 5 years and 70% of all estimates use less than 10 years of data.

Almost one third of the estimates use tax reforms that involve (among other things) the introduction of a top tax bracket and 46% of all tax reforms use (among other things) a reduction in the number of tax brackets. The mean unemployment rate is around 7.1 and ranges from 2.2 (=Netherlands in 2001) to 19.6 (=Poland in 2003). On average the Gini coefficient is around 30 and ranges from 20.9 (=Sweden in 2000) to 38 (=USA in 2005). The top 10% hold around 35% of national income, while the top 1% hold 11-12%. On average, the fraction of self-employed within a country is 10.724% and the share of modern taxes to GDP is 26.481%.

3 Meta-Regression Results

In subsection 3.1, I separately present the results for before (BD) and after deduction (AD) elasticities to account for different behavioural margins and quantify the influence of the estimation technique and sample restrictions on estimated elasticities. Then I add contextual variables to the regression in order to show how an estimate is correlated with non-controllable and given factors (see section 3.2).[14]

3.1 Baseline Results

To account for the range of behavioural responses, I run specification (4) on the *before* and *after* deduction subsample separately and present the results in Table 2 and 3. I define the most commonly used characteristic as a reference category and omit this feature such that reported coefficients need to be interpreted as a deviation from a particular characteristic to the corresponding reference category. In Table 2 and 3 reference categories are written in bold.[15]

[14]To verify the robustness of the baseline results, I apply various estimation techniques and further limit the dataset along certain dimensions (see section F in the (online) appendix).

[15]To facilitate the interpretation of my results, I provide an example: In Table 2, col 1: the constant is 0.072. Therefore, an average ETI of 0.072 is estimated with a specification that applies (a) mechanical tax rate

Table 1: Descriptive Statistics: Sources of Heterogeneity

	Before Deductions (BD) (N=852)		After Deductions (AD) (N=596)	
	Mean	Std. Dev.	Mean	Std. Dev.
Estimation Techniques				
Regression technique				
IV: mechanical tax rate changes	0.623	0.485	0.576	0.495
IV: (lagged) mechanical tax rate changes	0.038	0.19	0.154	0.362
IV: other	0.103	0.305	0.143	0.35
DID and IV	0.208	0.406	0.059	0.235
classic DID	0.028	0.166	0.069	0.253
Income Control				
Auten Carroll (1999)	0.299	0.458	0.193	0.395
none	0.228	0.42	0.282	0.45
Gruber Saez (2002) spline	0.163	0.37	0.146	0.353
Kopczuk (2005) type	0.224	0.417	0.357	0.48
other	0.086	0.28	0.022	0.146
Difference Length				
3 years	0.383	0.486	0.493	0.5
1 year	0.399	0.49	0.31	0.463
2 years	0.067	0.25	0.107	0.31
4+ years	0.151	0.359	0.089	0.285
Weighted by Income	0.442	0.497	0.326	0.469
Sample Restrictions				
Age Cutoff	0.462	0.499	0.475	0.5
Income Cutoff				
0-10k	0.286	0.452	0.171	0.377
none	0.112	0.315	0.153	0.36
10k-12k	0.21	0.408	0.346	0.476
12-31k	0.208	0.406	0.139	0.347
$>$ 31k	0.184	0.388	0.191	0.394
Variations across Countries and Time				
Country Group				
USA	0.548	0.498	0.673	0.47
Scandinavia	0.203	0.403	0.122	0.328
other countries	0.249	0.433	0.205	0.404
Mean year in study data	1993.452	7.444	1993.396	7.726
Estimation decade				
$<$ 1999	0.319	0.466	0.341	0.474
1990 - 2000	0.434	0.496	0.342	0.475
$>$ 2000	0.246	0.431	0.317	0.466
Publication Characteristics				
Publication decade				
2001-2010	0.405	0.491	0.542	0.499
$<=$ 2000	0.069	0.254	0.044	0.204
$>$ 2011	0.526	0.5	0.414	0.493
Published Type				
published in peer reviewed journal	0.741	0.439	0.721	0.449
working paper	0.259	0.439	0.279	0.449
Mean Year of Publication	2009.829	0	2009.829	0
Contextual Variables				
intro top bracket	0.306	0.461	0.285	0.452
reduce brackets	0.459	0.499	0.466	0.499
unemployment rate	7.098	2.813	7.135	1.387
Gini	30.85	5.322	31.954	4.931
top 10% inc. share	0.345	0.054	0.362	0.053
top 1% inc. share	0.112	0.036	0.125	0.034
fraction self-employed	10.699	3.823	9.629	2.079
share of modern taxes	26.311	9.438	24.552	9.022

Note: I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic. Reference categories are given in italics. More details can be found in the online appendix C.2. For a given estimate, contextual variables are merged via country and mean year of observation.

I gradually add the defined characteristics. In column (1) and (2) I only control for estimation technique, and in column (3) I account for sample restrictions. Results on country group coefficients are presented in column (4) and (5), with column (4) accounting for (estimation) decade, column (5) controlling for (publication) decade. Column (6) presents the most comprehensive specification.[16] As expected, I see that estimates that allow for deduction responses mostly reveal a larger constant and, therefore, are statistically more elastic to marginal tax rate changes compared to results obtained based on the before (BD) subsample. Contextual characteristics will be analysed separately and I show that mostly BD elasticities are positively correlated with tax- and economy related characteristics except for measures of inequality. In the following I simultaneously present obtained results for both subsamples by source of heterogeneity.

Estimation techniques. The results reveal huge differences in terms of *regression techniques*, *income control*, *difference length* and *weighting by income* in both subsamples of elasticities. The reference specification in column (1) is defined as a specification that uses mechanical tax rate changes as an instrument, log base- year income control and a three-year difference length. For example, it refers to the most standard specification used by Kleven and Schultz (2014) in their baseline specifications. On average, such a specification yields a BD elasticity of 0.072 and an AD elasticity of 0.443. We observe a difference between BD and AD elasticities because an AD elasticity concept is exposed to larger income fluctuations and allow a wider range of behavioural responses (e.g. itemized deductions). There is also a mechanical effect because an AD tax base is smaller than a BD tax base and income changes have a relatively larger effect (Gruber and Saez, 2002). As expected elasticities that consider

changes as an IV, (b) uses log-base year income control (Auten Carroll) and (c) uses a difference length of three years. An average estimate that is based on (a) mechanical tax rate changes as an IV, (b′) an NO income control and (c) a difference length of three years (everything else equal) is equal to: 0.072-0.210=-0.138. If 'no restriction' defines the base category, it means that a particular estimate is not restricted with respect to a certain characteristic. For instance, the baseline category for age restriction is 'no restriction'. Hence, estimates need to be interpreted in reference to other estimates that do not apply an age cutoff.

[16] Multicollinearity might be a problem in the regressions resulting in standard errors that are too large. This makes it difficult to isolate the influence of a single variable from overall influence. Therefore, I check if the variance inflation index is below 10 such that the presented results are reliable within every estimation. Except for column (6) in Table 2 and Table 3 this condition holds.

a deduction component are more sensitive to the estimation technique.

As already noted in most primary studies, a regression that does not consider any *income control* leads to lower and often negative BD elasticities. Giertz (2007), however, provides evidence that the direction of bias in the case of no income controls depends on the underlying income dynamics and the direction of tax change. On average, estimates that do not use income controls lead to a decrease of 0.2 in the BD elasticity. This result is quite robust even in the most sophisticated specification as shown in column (6).

Most studies follow Auten and Carroll (1999) and include log base-year income as an explanatory variable. Interestingly, all other kinds of income control variables (in most cases more sophisticated ones) lower elasticities in both but in particular in the AD subsample compared to this approach. On the one hand, income controls capture a potential (upward or downward) bias induced by non-tax related factors (e.g. increasing inequality or mean reversion) but on the other hand they might absorb too much identifying variation (see Saez et al., 2012 for a discussion). It is worth highlighting that Kopczuk-type income controls lower AD elasticities (on average) by 0.380 compared to a log base-year income control while other types of income controls (mostly splines) also decrease AD elasticities but at a lower rate.

The results suggest that the chosen *difference length* has different effects on BD and AD elasticities. In the BD subsample, all specifications with a two-year time window have a significant lower elasticity compared to specifications based on three-year differences. It is reasonable to assume that BD estimates mainly reflect labour supply responses and that individuals do not easily adjust them in response to tax rate changes. Matikka (2016) has detailed information on wage rates, working hours, and deductions. He provides tentative evidence that both work effort (wage rates) and labour supply (working hours) are not responsive to tax rate changes. Turning to AD elasticities, it seems that one- or two-year differences significantly increase elasticities. Although taxpayers cannot easily adjust their labour supply, they can adjust their taxable income. There are opportunities to do so depending on the tax system in place. Either taxpayers change their spending on tax-deductible items, even claiming questionable deductions (see Paetzold, 2017 or Doerrenberg et al., 2017) or over-report their deductions (Kleven et al., 2011). Income shifting between periods as documented by Goolsbee (2000) might also drive the results.

If BD elasticities are *weighted by income*, they tend be marginally lower compared to unweighted estimates. This is in line with the literature. Contrary to what was found in the literature, weighted AD elasticities reveal a negative sign. Interestingly, almost every study either completely ignores weighting as a model choice or it weights every estimate. Gruber and Saez (2002) find that a weighted ETI is very similar to the unweighted ETI, while an unweighted gross elasticity is substantially lower than the weighted elasticity. They argue that most of the behavioural response comes from high gross-income taxpayers but not necessarily from those with a high taxable-income, because of the central role of itemization. Giertz (2010*b*) on the other hand finds that unweighted ETI estimates are smaller than income-weighted estimates. It is worth noting that weighting by income is a controversial model choice (Weber, 2014*b*).

Sample Restrictions. An *age cutoff* restricts income and employment fluctuations at the beginning and end of a person's working life. Such a cutoff has contrasting effects on elasticities depending on the subsample. In particular, estimates in the BD subsample are lowered when a primary study restricts its data to a certain age. In contrast to BD elasticities, I observe a stable positive coefficient for AD elasticities. Most interestingly, *income cutoffs* have no effect on estimated BD elasticities. This is in stark contrast to findings for the AD subsample where an income cutoff and its value matters greatly. Both cutoffs have a large effect on AD elasticities, while the effect on BD elasticities is negligible and often insignificant and close to zero. This is an interesting finding since it is unclear whether or not a certain cutoff (and its level) helps or impairs a correct identification.

Table 2: WLS before deductions baseline results

Dependent Variable: Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.050*	0.051*	0.035	0.007	0.013	0.018
	(0.025)	(0.028)	(0.024)	(0.015)	(0.008)	(0.013)
IV-other	0.067	0.069	0.073	0.069	0.054	0.042
	(0.063)	(0.058)	(0.056)	(0.060)	(0.054)	(0.057)
DID-IV	0.289***	0.259***	0.291***	0.262***	0.295***	0.251***
	(0.061)	(0.075)	(0.083)	(0.050)	(0.066)	(0.057)
DID-classic	0.321***	0.292***	0.159**	0.213***	0.137*	0.094
	(0.070)	(0.082)	(0.076)	(0.058)	(0.069)	(0.071)
Income Control (omitted: Auten Carroll)						
none	-0.210***	-0.212***	-0.207***	-0.206***	-0.204***	-0.207***
	(0.027)	(0.025)	(0.029)	(0.030)	(0.031)	(0.029)
Gruber Saez Spline	-0.018***	-0.013*	-0.012*	-0.008	-0.009	-0.005
	(0.005)	(0.006)	(0.006)	(0.009)	(0.008)	(0.012)
Kopczuk	-0.016**	-0.014**	-0.007	-0.008	-0.005	-0.005
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.009)
other	-0.034*	-0.073	-0.008	-0.012	-0.001	-0.029
	(0.020)	(0.044)	(0.010)	(0.014)	(0.008)	(0.018)
Difference Length (omitted: 3-years)						
1 year	0.069	0.061	0.041	0.038	0.032	0.008
	(0.074)	(0.066)	(0.058)	(0.052)	(0.053)	(0.035)
2 years	-0.010	-0.010	-0.043***	-0.049**	-0.057***	-0.071**
	(0.024)	(0.025)	(0.014)	(0.019)	(0.019)	(0.030)
4 years and more	0.079	0.081	-0.003	-0.015	-0.019	-0.015
	(0.054)	(0.055)	(0.016)	(0.021)	(0.025)	(0.021)
Weighting by Income (omitted: no restriction)						
Weighting by Income applied		-0.045*				-0.089*
		(0.026)				(0.048)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied			-0.190***		-0.070	-0.100
			(0.051)		(0.072)	(0.075)
Income Cutoff applied (omitted: 0-10k)						
none			-0.025		-0.006	-0.047
			(0.020)		(0.015)	(0.041)
10k-12k			-0.010		-0.014	-0.021*
			(0.011)		(0.010)	(0.013)
12k-31k			0.010		0.006	0.009
			(0.007)		(0.008)	(0.015)
>31k			0.009		0.001	0.003
			(0.020)		(0.014)	(0.016)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia				-0.108***	0.008	0.049
				(0.032)	(0.080)	(0.072)
other countries				0.087	0.155**	0.229***
				(0.063)	(0.072)	(0.084)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.285*	0.278*
					(0.157)	(0.148)
after 2010					-0.164**	-0.117
					(0.067)	(0.079)
(Estimation) Decade (omitted: 1980s)						
1990s				-0.046***		-0.046***
				(0.002)		(0.004)
2000s				-0.030***		-0.096**
				(0.011)		(0.042)
Constant	0.072***	0.111***	0.252***	0.207***	0.285***	0.352***
	(0.006)	(0.031)	(0.055)	(0.034)	(0.052)	(0.064)
Observations	852	852	852	852	852	852
Adjusted R^2	0.573	0.585	0.636	0.672	0.660	0.695

Note: Columns (1) to (7) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: WLS after deductions baseline results

Dependent Variable: Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.447***	0.375***	0.480**	0.408***	0.328**	0.307***
	(0.129)	(0.085)	(0.224)	(0.137)	(0.142)	(0.106)
IV-other	-0.358***	-0.350***	-0.314***	0.174	0.031	0.101
	(0.121)	(0.129)	(0.096)	(0.276)	(0.175)	(0.230)
DID-IV	-0.677***	-0.689***	-0.954***	-0.252	-0.545	-0.514
	(0.178)	(0.175)	(0.228)	(0.364)	(0.351)	(0.415)
DID-classic	0.129	0.124	0.051	0.279**	0.204***	0.159***
	(0.114)	(0.110)	(0.105)	(0.127)	(0.075)	(0.046)
Income Control (omitted: Auten Carroll)						
none	0.128	0.129	-0.041	0.068	-0.173*	-0.137
	(0.126)	(0.127)	(0.100)	(0.140)	(0.092)	(0.095)
Gruber Saez Spline	-0.033	-0.006	-0.093	-0.065	-0.185	-0.142
	(0.105)	(0.128)	(0.107)	(0.130)	(0.109)	(0.135)
Kopczuk	-0.380***	-0.131	-0.471***	-0.248**	-0.150	0.021
	(0.114)	(0.166)	(0.107)	(0.107)	(0.142)	(0.163)
other	-0.327*	-0.198	-0.568***	-0.177	-0.238**	-0.190
	(0.180)	(0.210)	(0.147)	(0.152)	(0.113)	(0.142)
Difference Length (omitted: 3-years)						
1 year	0.026	0.009	0.193***	0.015	0.156**	0.161***
	(0.133)	(0.148)	(0.045)	(0.137)	(0.061)	(0.058)
2 years	0.199**	0.217**	0.088	-0.054	0.010	-0.028
	(0.077)	(0.080)	(0.094)	(0.182)	(0.083)	(0.119)
4 years and more	0.258	0.150	-0.027	0.096	-0.315	-0.468*
	(0.171)	(0.166)	(0.200)	(0.199)	(0.205)	(0.271)
Weighting by Income (omitted: no restriction)						
Weighting by Income applied		-0.260***				-0.252*
		(0.078)				(0.142)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied			0.504***		0.468***	0.503***
			(0.077)		(0.120)	(0.112)
Income Cutoff applied (omitted: 0-10k)						
none			0.359***		0.451***	0.437***
			(0.062)		(0.060)	(0.049)
10k-12k			0.255**		0.586**	0.739***
			(0.114)		(0.232)	(0.230)
12k-31k			0.141**		0.028	0.117
			(0.066)		(0.077)	(0.085)
>31k			1.061***		1.069***	1.029***
			(0.262)		(0.240)	(0.205)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia				0.013	0.207	0.253
				(0.109)	(0.237)	(0.209)
other countries				0.578**	0.523**	0.614***
				(0.256)	(0.208)	(0.194)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.818	0.815
					(0.628)	(0.660)
after 2010					-0.160	-0.147
					(0.145)	(0.126)
(Estimation) Decade (omitted: 1980s)						
1990s				-0.041		-0.070***
				(0.036)		(0.009)
2000s				-0.400		-0.198
				(0.268)		(0.164)
Constant	0.443***	0.451***	0.030	0.325**	-0.308	-0.273
	(0.113)	(0.108)	(0.091)	(0.131)	(0.244)	(0.211)
Observations	596	596	596	596	596	596
Adjusted R^2	0.580	0.588	0.729	0.618	0.773	0.785

Note: Columns (1) to (7) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Variations across time and countries. Column (4) and (5) take into account *country group*. While column (4) controls for *estimation decade*, column (5) shows the results for *publication decade*. Compared with estimates for the US, results for other countries are larger. Estimates published prior to 2001 are larger or smaller if they are reported after 2010 compared to estimates derived from studies reported between 2001 and 2010.

For the US Giertz (2007) shows that an ETI for the 1990s is smaller compared to an estimate derived for the 1980s. My analysis is able to generalize this finding across the entire empirical literature. Estimates derived in the 1980s are always larger than those derived for the 1990s or 2000s. Most of the previous findings prevail in both subsamples. Column (6) shows the results of the most comprehensive specification that accounts for all the defined sources of heterogeneity. Unfortunately, multicollinearity seems to influence the results to the extent that the precision of some coefficients vanishes.

3.2 Contextual Factors

The following exercise allows a descriptive analysis on how economy- and tax reform related factors are linked to an elasticity. Results are displayed in Table 4. The baseline specification involves controls for estimation technique, income controls and difference length and I gradually take into account contextual factors.[17]

I ignore all other tax-related issues (e.g. base broadening) that might have been occurring simultaneously. Given that wealthier people tend to be more responsive, I expect a positive relationship between an *introduction of a top tax bracket* and behavioural responses. Contrary to expectation, the coefficient of an introduction of a top tax bracket is insignificant and close to zero.

There is a positive correlation between *inequality measures* and elasticities. In particular, AD elasticities are highly correlated with top income shares. All studies use data that covers years where mostly tax cuts for high-income taxpayers have been implemented. As Alvaredo et al. (2013) observe, there has been a widening of the income distribution and top

[17] I only show the relevant coefficients and full results can be found in the (online) appendix.

Table 4: WLS: Contextual Factors

Dependent Variable: Elasticity:	Before Deduct.	After Deduct.
Additional Variables		
Intro top bracket	-0.032	-0.086
	(0.093)	(0.116)
Gini Coefficient	0.010***	0.002
	(0.003)	(0.012)
Top 10%	0.832***	3.301***
	0.280	(1.199)
Top 1%	0.794**	3.645**
	0.361	(1.689)
Unemployment Rates	-0.009**	0.049
	(0.004)	(0.080)
Fraction of self-employed	0.018**	0.002
	(0.008)	(0.034)
Modern taxes (in 2005)	-0.010***	0.013
	(0.002)	(0.009)

Note: Both columns are estimated using Weighted Least Squares with precision as weights. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The baseline specification only includes controls for estimation technique (regression technique, income control and difference length) and I gradually add each contextual characteristic separately. Full results can be found in the (online) appendix. For the first two characteristics, I compare the first and last year of a data period. Remaining characteristics are merged via mean year of observation. For observations that are based on a classic DID approach, I do not have information of the share of self-employed people that corresponds to the respective mean year of observation.

tax rates have moved in the opposite direction from top pre-tax income shares. While top pre-tax income shares are rising, top tax rates are decreasing. This confirms the fact that not only current but also past tax policy still has an effect on estimated elasticities and that the underlying context matters when interpreting elasticities. Moreover, it might be the case that income control variables do not fully account for such a development and this leads to an upward bias of AD elasticities.

Business cycle effects are related to BD elasticities. There is a negative relationship between *unemployment rate* and BD elasticities but I do not find any correlation with AD elasticities.

As shown by Kleven et al. (2016) and Kleven et al. (2011), there is a close relationship between tax enforcement, tax compliance and third party information reporting. The regression results show that third party information reporting is negatively related to BD elasticities. Given that self-employed people have greater control over their income, there is a positive correlation between BD elasticities and the share of self-employed workers in an economy. The share of modern taxes is negatively related to BD elasticities. Neither measure influences AD elasticities. This strengthens the fact that AD responses are mainly driven by avoidance behaviour. Most taxpayers respond via itemized deductions that are not subject to third party information reporting.

It is evident from the results that behavioural elasticities are also endogenous with respect to contextual factors. More precisely, behavioural responses captured by a BD and AD elasticity are also correlated with external factors. This highlights the fact that elasticities are potentially influenced by past and current circumstances.

4 Selective Reporting Bias

In the last part of the analysis, I check for the presence of a selective reporting bias. Publishing statistical results that reject the hypothesis of no effect reflects a general desire to report findings that are supposed to be trustworthy. Moreover, researchers naturally want to publish results that exhibit intuitive magnitudes. Publication or reporting selection bias has been identified in other areas of empirical work. Ashenfelter et al. (1999) review the literature on the rate of return on schooling investment and show reporting selection bias in favour of significant and positive returns to education. Card and Krueger (1995) find such biases in the minimum wage literature and Lichter et al. (2015) in the literature on labour demand elasticities. A recent study by Brodeur et al. (2016) uses more than 50,000 tests published in three top economic journals and find that researchers are prone to choose more 'significant' specifications in order to increase the chance of publication. Moreover, they show that scientists use z-statistics of 1.64 or 1.96 as reference points.

I provide evidence of selective reporting bias in section 4.1. To start the analysis, I follow Brodeur et al. (2016) and plot the distribution of z-statistics and, I then examine the relationship between standard errors and estimates (Card and Krueger, 1995). Finally, I check statistically whether publication bias is prevalent. In section 4.2, I discuss how biased estimates can lead to skewed predictions of optimal tax rates. In addition, I examine the role of administrative datasets. For future research, I present potential improvements that might reduce this type of bias.

4.1 Evidence

Distribution of z-statistics An obvious type of bias is the excessive production and selection of significant results. Given that z-statistic = beta coefficient / standard error, there are three ways to receive significant values. First, to find a specification where standard errors are low enough. Second, to search for a specification where coefficients are large enough to offset 'large' standard errors. Or third, through a combination of these two things. Since research on behavioural responses to taxation relies on administrative datasets with a large number of observations, standard errors are generally small.

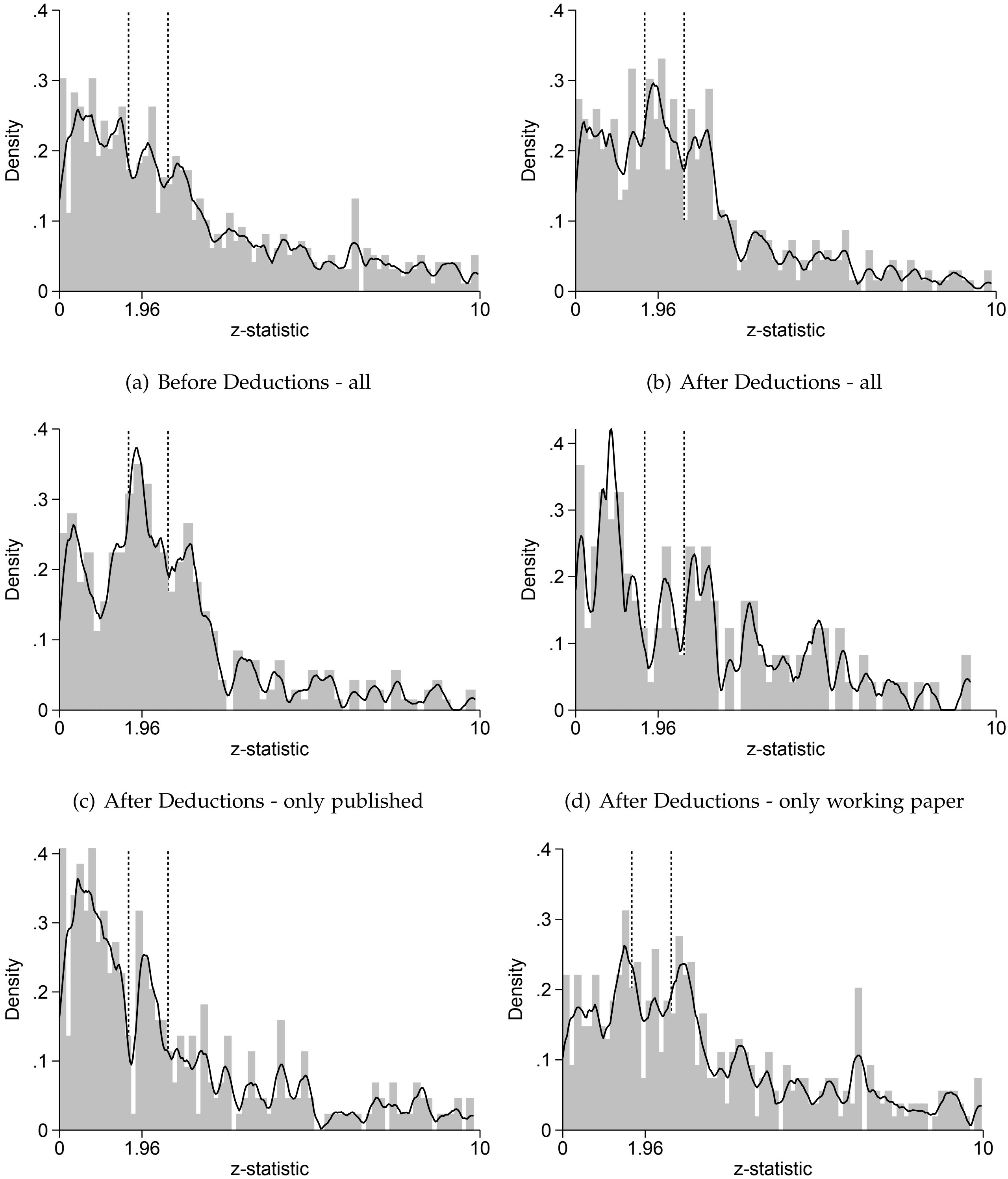
I plot the distribution of z-statistics for the two subsamples (see Figure 2).[18] Subfigure (a) shows the BD and Subfigure (b) the AD subsample. In accordance with Brodeur et al. (2016), I observe a local maximum around 2 (= 5% significance) and also a valley before this. Moreover, I also observe a spike around 1.64 (= 10% significance) and around 3 (= 0.05-0.01% significance). These simple graphs provide evidence consistent with the existence of 'p-hacking'.[19]

Interestingly, I observe that this pattern is more pronounced in the AD subsample. A reason might be that for a long time it has been thought that only AD (more precisely taxable income) elasticities matter for policy analysis. Hence, a significant elasticity of taxable income was much more important in terms of credibility and publication than other significant elasticities. This fact is confirmed by Subfigure (c) and (d). Here, I divide the AD subsample into estimates reported in journal articles and working papers. Among published estimates, the maximum around 2 is even more pronounced. It is unclear whether a researcher chooses the most credible findings in the first place to increase the chances of

[18] I formally tested the equality across the distributions. I applied a Kolmogorov-Smirnov test which tests whether different t-distributions are equal. In all three cases, I am able to reject the hypothesis that this is the case.

[19] There are other peaks and valleys across the distributions. Unlike Brodeur et al. (2016) I use considerably fewer observations, with the result that my graphs appear to be more 'bumpy'.

Figure 2: Raw distribution of z-statistics



(a) Before Deductions - all

(b) After Deductions - all

(c) After Deductions - only published

(d) After Deductions - only working paper

(e) Before Deductions - prior to Chetty (2009)

(f) Before Deductions - after Chetty (2009)

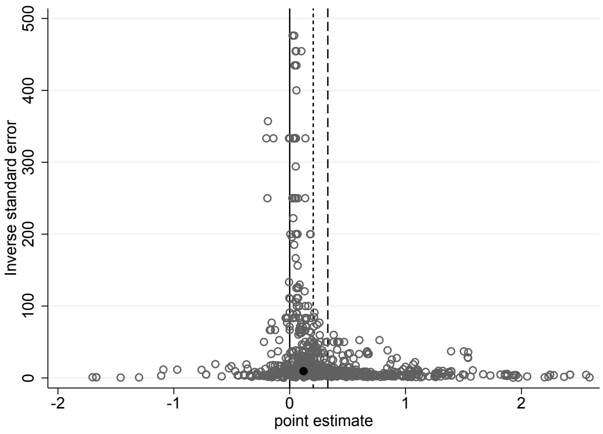
Note: All graphs plot the distribution of z-statistics and the significance level of 5% (1.96) and also the band-width from 10% to 1% are highlighted. Subfigure (a) plots all estimates from the Before Deductions (BD) subsample and Subfigure (b) for the After Deductions (AD) subsample. Subfigures (c) and (d) split the AD subsample into estimates published in journals and estimates reported in working papers. Subfigures (e) and (f) split the BD subsample into estimates that are published prior to and after 2009.

publication and/or that referees/journals prefer significant estimates. Moreover, journal editors often require authors to streamline their papers prior to publication, leading them to limit the number of tables and figures in their paper. Therefore, it is unclear who chooses which estimates are published.

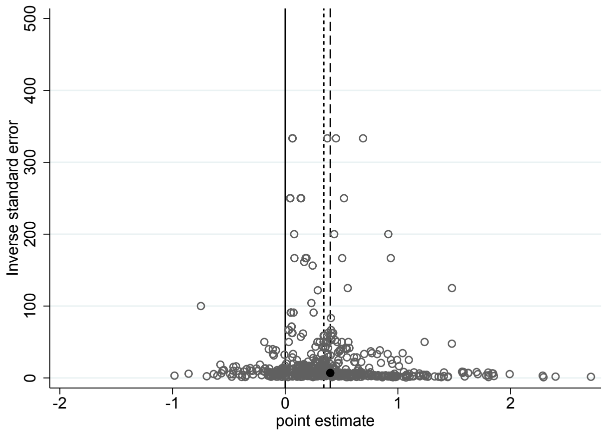
Chetty (2009) shows that the excess burden of taxation depends on a weighted average of taxable income and total earned income elasticities. Since the publication of his study in 2009, BD (e.g. gross income) elasticities have begun to receive more attention than ever before and people have started to think more carefully about both types of elasticities. Therefore, I divide the BD subsample into estimates reported prior to and after 2009. As seen in Subfigure (e), I observe a larger insignificant mass before 2009 and a huge spike at 1.96 (=5% significance level) and a missing mass before. After 2009 I observe a much smaller insignificant mass but still a spike at 1.64 (=10% significance), 1.96 (=5% significance) and now also around 3 (=0.05-0.01% significance level). The graphical evidence confirms that the share of significant BD elasticities has increased over time.

Relationship between estimate and standard error. In a second step, I follow Card and Krueger (1995) and analyse the relationship between an estimate and its standard error obtained in different studies. I apply a standard procedure and use what is known as a funnel plot in order to analyse the correlation. Funnel plots are simple scatter plots of elasticity estimates on the horizontal axis and their precision (=inverse of standard error) on the vertical axis. The most precise estimates are close to the top of the funnel and as precision decreases, the dispersion of estimates increases. The shape of the graph should look like an inverted funnel. In the absence of selective reporting bias, there should be no systematic relationship between estimates and standard errors. All imprecise estimates should have the same probability of being reported. The funnel should be symmetric with the estimates randomly distributed around the population elasticity. If the estimates are correlated with their standard errors, the funnel can take an asymmetric shape. This might happen when researchers select only significant estimates and/or estimates with a certain

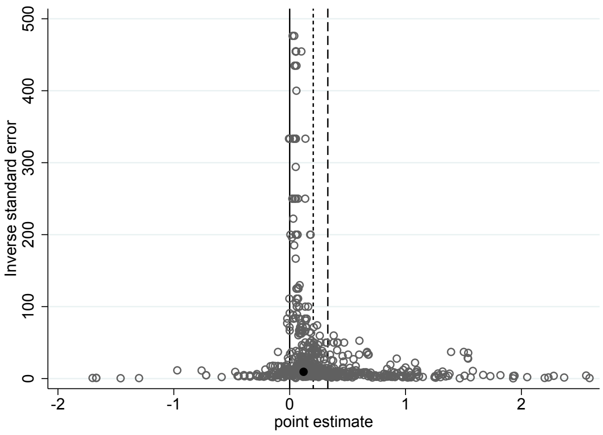
Figure 3: Funnel Plot



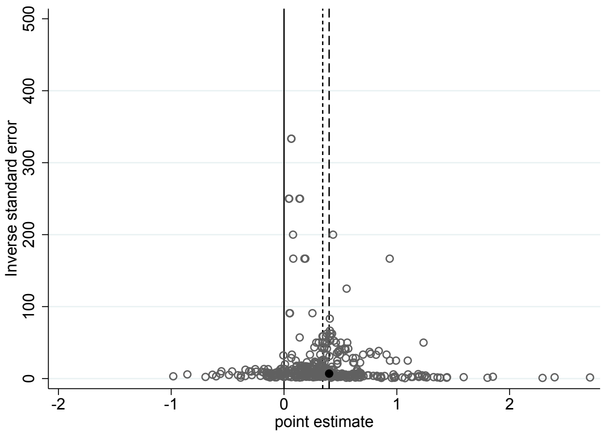
(a) Before Deductions - all



(b) After Deductions - all



(c) Before Deductions - only income control(s)



(d) After Deductions - only income control(s)

Note: Funnel plots are presented separately for the before and after deductions subsamples. The short dashed line denotes the median and the dashed line the mean of the corresponding (full) subsample. In the dataset the median (mean) BD elasticity is 0.203 (0.300) and 0.343 (0.389) respectively for an elasticity that considers deductions. The base results from Gruber and Saez (2002) are highlighted in black. They report coefficients of 0.4 with a standard error of 0.144 for the ETI and 0.12 with a standard error of 0.106 for the elasticity of broad (=gross) income. Subfigures (a) and (b) display all collected estimates. Subfigures (c) and (d) are based on a subset of estimates that rely on a specification with income control(s).

sign (e.g. omit negative values) such that their results are consistent with theory.[20]

Figure 3 plots BD and AD elasticities separately along with their precision. I highlight the mean and median as well as estimates obtained by Gruber and Saez (2002). Subfigures (a) and (b) are based on the full sample of estimates, while I restrict the sample to estimates that

[20] As well as a graphical analysis, I formally checked for funnel asymmetry and conducted a so-called Funnel-asymmetry test as proposed by Egger et al. (1997). In all cases, I am able to reject the hypothesis of funnel symmetry. Besides selective reporting bias, there are other reasons why funnel asymmetry could arise (e.g. data irregularities or low methodological quality of some studies). See Sterne et al. (2004).

rely on income controls and therefore explicitly account for non-tax related income growth in Subfigure (c) and (d). Subfigures for BD and AD reveal some noticeable differences. First, I observe a more pronounced missing mass on the negative side in the BD compared with the AD subsample. According to theory an increase in the marginal tax rate lowers the net of tax rate, which in turn should reduce taxable income in the simplest case with no income effects or frictions. If a researcher receives a negative value, this translates into a situation where the government can tax income by 100% while the people earn/work even more. Hence, it seems plausible that researchers tend to put more trust in positive results to keep in line with theory. This behaviour causes a positive relationship between standard errors and estimates. AD elasticities allow a wider range of responses and it is also well-documented that running the exact same specification results in a larger AD elasticity compared to an BD elasticity (Gruber and Saez, 2002). The chance of reporting negative values is therefore larger for an elasticity that does not consider deductions. This might explain why I observe a larger missing mass on the negative side in the BD subsample.

Within the AD subsample, it appears that researchers tend to report an estimate between 0 and 0.4 with a higher probability compared to estimates ranging from e.g. 0.4 to 0.8. I expect a negative relationship between standard errors and estimates and therefore a downward bias of AD estimates.

Distribution of estimates. Another kind of selection reporting bias arises, if researchers use well-known results as a reference point and hence are inclined to report only results that are in line with these findings. Piketty and Saez (2013) write in their handbook chapter that an elasticity of 0.25 seems realistic (same as Chetty, 2009), 0.5 is high and 1 is extreme. As seen in Figure 1, there is a general tendency to report results that lie within an interval of 0 and 1. I observe a considerable excess mass between 0.7 and 1. This indicates an aversion to report a value above 1. In their well-known and widely-cited survey, (Saez et al., 2012, p. 42) refer to their estimates and write '[...]. While there are no truly convincing estimates of the long-run elasticity, the best available estimates range from 0.12 to 0.4. [...]' and '[...]

0.25 corresponds to the mid-range of estimates found in the literature. [. . .]′ With regard to the AD-funnel, there is a slight incline to report values between 0 and 0.4 (=mean of AD estimates in the dataset). Saez et al. (2012) and Gruber and Saez (2002) are the most widely-cited articles in the literature on taxable income elasticities, and results obtained by Gruber and Saez (2002) might serve as another benchmark. They report baseline results of 0.4 for the elasticity of taxable income with a standard error of 0.144 and 0.12 with a standard error of 0.106 for the elasticity of broad (gross) income. A simple descriptive analysis shows that estimates close to the mentioned values are indeed reported more often. Within the AD subsample 34.8% of all estimates lie within one standard deviation around Gruber Saez's ETI of 0.4. It looks similar within the BD subsample, where 27.32% of all estimates lie within one standard deviation close to their broad income elasticity of 0.12.

Table 5: Testing for Selective Reporting Bias: Before Deductions elasticities

Dependent Variable:	BD	BD	BD	BD	AD	AD	AD	AD
Elasticity:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Error	3.725***	4.203***	0.768	0.679	-0.226	-0.926***	-0.494	-0.470
	(0.787)	(0.580)	(0.580)	(1.034)	(0.217)	(0.337)	(0.434)	(0.572)
Journal impact factor		-0.013				0.038**		
		(0.009)				(0.015)		
Std.Error* Impact Factor		-0.051				0.074***		
		(0.037)				(0.019)		
Dummy if obs > median(obs)			0.816***				-0.258	
			(0.259)				(0.284)	
Std.Error*D if obs > median(obs)			4.467***				0.074	
			(1.128)				(0.582)	
Dummy reported prior to 2009				0.632**				-0.118
				(0.309)				(0.350)
Std.Error*D reported prior to 2009				3.839***				0.128
				(1.454)				(0.681)
Constant	0.921***	1.044***	0.449***	0.469**	0.263	-0.212	0.323	0.207
	(0.180)	(0.213)	(0.127)	(0.198)	(0.168)	(0.283)	(0.202)	(0.325)
Observations	852	852	852	852	596	596	596	596
Adjusted R^2	0.618	0.631	0.645	0.630	0.580	0.604	0.596	0.590

Note: Columns (1) to (8) are estimated using Weighted Least Squares using precision as weights. I control for estimation technique (= regression technique, income control and difference length. Full results can be found in the (online) appendix in Tables 21 and 22. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included standard errors as explanatory variables are normalized such that they can be interpreted as a standard deviation.

Regression results. To statistically examine the presence of selective reporting bias, I take the specification of column (1) of Table 2 and respectively Table 3 as the baseline specification (= WLS with estimation technique controls).[21] Point estimates and respective

[21] As a robustness check, I also apply wild-cluster bootstrap methods instead of WLS and my results barely change.

standard errors should be independent according to random sampling theory (Card and Krueger, 1995; Stanley and Doucouliagos, 2010). Therefore, I explicitly control for an estimate's standard error and subsequently add more publication- related characteristics. For the sake of interpretation, I normalize the standard error. I control for publication type and also for a simple impact factor.[22] I also check if the number of observations used in primary estimations influence the results. I calculate the median of observations for each subsample and create a dummy variable if an estimate is based on a dataset that is smaller or larger compared to the median sample size of all other collected estimates. Studies released after Chetty (2009) might be correlated differently, and I therefore include a dummy variable indicating if an estimate is reported after 2009. Overall, the regression results confirm what can already be seen in the figures presented before.[23]

The standard error has a strong and statistically significant effect on BD elasticities (column (1)). As already shown, the funnel plot for BD estimates indicates significant selective reporting bias in the estimates towards more positive elasticities. In columns (3) and (7) I account for the fact that larger datasets increase the chance of yielding standard errors that are small enough to produce significant and trustworthy results. Only for BD elasticities, the relationship is significantly positive. In columns (4) and (8) I also include a dummy variable indicating if an estimate was reported prior to Chetty (2009). Both aspects influence BD elasticities but not AD elasticities.

While the graphical evidence and regression results indicate an upward reporting bias among BD estimates, the reporting bias goes in both directions in the AD subsample but the downward bias appears to be more dominant. As seen in figure 3, there is an aversion to reporting negative estimates and estimates above 0.4 and in particular above 1. It appears that selective reporting bias is more prevalent in journals of a high quality (compare column 6 in Table 5). This is line with what distributions of z-statistics among AD elasticities have shown. It is more pronounced at reference points related to statistical significance such as

[22]I downloaded the IDEAS RePEc simple impact factor (22.06.2016) and working papers receive a value of 0.
[23]Obviously, publication year and number of observations are correlated with the standard error. Therefore, I checked the (mean) variance inflation factor, which is below 10.

1.96 compared to the BD subsample and this pattern is more visible in published articles compared to working papers (see Figure 2).

4.2 Discussion

Tax Policy Effects. Selective reporting bias can lead to skewed predictions with regard to optimal tax rates and the welfare cost of taxation. To illustrate these effects, I derive and compare optimal marginal top tax rates for different elasticity values. There are three key elements that determine an optimal marginal top income tax rate: (1) the elasticity of taxable income, (2) the thickness of the top tail of the income distribution described by a pareto parameter a and (3) the redistributive taste of a government defined by g (Mirrlees, 1971; Diamond, 1998; Saez, 2001). The optimal marginal top tax rate is defined as $t = \frac{1-g}{1-g+a*e}$.[24] If we assume that the social marginal welfare weight on top income earners, g, is equal to zero, the derived tax rate is also called the revenue-maximizing tax rate. This defines an upper limit of a tax rate a government can set before it starts to lose revenue.

Table 6: Optimal marginal top tax rates

Elasticity	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Panel A: $g = 0$											
	1	0.83	0.71	0.625	0.56	0.5	0.45	0.42	0.38	0.36	0.33
Panel B: $g = 0.15$											
	1	0.81	0.68	0.58	0.52	0.46	0.41	0.38	0.35	0.32	0.32

Note: Optimal rates are computed using the formula defined as $t = \frac{1-g}{1-g+a*e}$. The pareto parameter is set equal to 2.

As I have shown, it is difficult to summarize selective reporting bias into a single number. Therefore, I provide suggestive evidence on the resulting effects. A large variation in elasticity estimates translates into a broad range of calculated tax rates (see Table 6). My illustration relies on plausible values for g and a (see Diamond and Saez, 2011).[25] An aversion against elasticity estimates above 0.4, translates into an aversion against revenue

[24] I consider a utility function that abstracts from any income effects (Diamond, 1998; Saez, 2001). I also ignore other tax system characteristics and extensions with respect to the model of optimal income taxation.

[25] It should be noted that pareto parameters and marginal social welfare weights differ across countries and years. I ignore these aspects for simplicity.

maximizing tax rates below 0.56. Given that reported elasticities are barely above 1, an optimal tax rate below 0.32 can hardly be predicted. This simple analysis shows that selective reporting bias has severe effects on policy analysis.

Role of Administrative Data. Unlike the literature on the effects of taxation on labour supply, which relies mostly on survey data, the ETI-literature exclusively uses administrative data. Administrative tax-return data offers precise information about a tax unit's income situation and the data is collected by tax agencies. Therefore, the measurement error is much smaller compared to survey data. However, a replication of findings is in most cases impossible. Researchers may disclose their programming files but they are not allowed to release their dataset for confidentiality reasons. In many cases data access is even restricted to a small number of people and/or it is very time-consuming and costly.

Using such datasets can be costly because of the following reasons. First, in some cases, administrative tax return data is not available via remote access and it requires a researcher to work on-site with the data. Second, the data is often available only in the respective native language. Researchers therefore not only need to have in-depth knowledge about the underlying and respective tax system but also knowledge of the respective language. These reasons are particularly relevant for non-US countries. Third, the number of data users is restricted in some countries because of limited capacity at the respective statistical office. In summary, the use of administrative tax records is associated with a large amount of fixed costs and administrative barriers that makes replicating results very difficult.

As mentioned by Slemrod (2016), researchers should be encouraged to provide as much information as possible to enhance at least a comprehensive understanding of the calculations behind the elasticities they obtain. Since reporting bias influences elasticities, possible solutions might also involve reporting standards or even a pre-analysis plan (see Burlig, 2018; Christensen and Miguel, 2016). Potential reporting standards might involve information about data irregularities and more details about the constant (and therefore artificial) tax base and the tax reform used for identification. A pre-analysis plan might cover the study design, outcome measures, details about subgroup analysis and sample restrictions. As a result researchers may not pick the most suitable estimate in the first place, but rather reveal underlying sensitivity.

5 Conclusion

My study applies meta-techniques to identify and to assess different explanations for the varying sizes of estimated elasticities. The magnitude of such estimates is of major importance for tax policy analysis. I differentiate between real responses (before deduction elasticities) and avoidance behaviour (after deduction elasticities) and use 1,448 estimates from 51 studies.

My meta-regression analysis generalizes findings across the entire empirical literature and confirms previous results. Elasticities that include deductions are more elastic compared to elasticities that are based on gross income. Both types of elasticities are sensitive with respect to the underlying empirical strategy and sample restrictions. Plus, they vary across country and time. I provide new evidence on the relationship between elasticities and contextual variables. Current and also past tax policy, income inequality, business cycle and third party information reporting requirements can affect existing elasticity estimates.

Apart from this, my study shows that selective reporting bias is prevalent in the literature of taxable income elasticities. There is an upward reporting bias among BD elasticities while the reporting bias for AD elasticities goes in both directions with a downward bias appearing to be dominate. Both measures are more likely to get published if they lie in a range between 0 and 0.4. Researchers tend to report behavioural responses that are in line with existing theory and also use existing evidence and significance levels as reference points. In general, such behaviour is more pronounced among AD compared to BD elasticities. Since the contribution of Chetty (2009) showing that not only AD but also BD elasticities matter for welfare analysis, reported BD elasticities have begun to receive more attention over time and therefore have become more significant. I illustrate how biased elasticities can lead to skewed predictions regarding optimal top tax rates under reasonable assumptions.

Unlike the literature on the effects of taxation on labour supply, which relies mostly on survey data, the ETI-literature exclusively uses administrative tax-return data. On the one hand, administrative tax-return data provides precise information about a tax unit's income situation that is needed for estimation but, on the other hand, data access is often restricted to a small number of people and its utilisation is costly in various dimensions. Researchers should be encouraged to provide as much information as possible to promote a comprehensive understanding of the obtained elasticities. Reporting standards or even a pre-analysis plan might reduce the problem of selective reporting bias.

Following Kopczuk (2015) 'sensitivity of reported estimates is due to model mis-specification, lack of credible variation and poor understanding of the data'. I agree and argue that future research should pay more attention to income dynamics and the tax reform used for identification. For instance, questions like who is affected and to what extent income grows irrespective of the reform should be analysed more carefully. Based on this, one may find a strategy that produce robust results. I argue that not only more descriptive evidence might be helpful, but more transparency with respect to the underlying tax simulation model is essential. Moreover, instead of proving a single estimate, a range of estimates is more meaningful in order to shed light into the heterogeneity of behavioural responses across the income distribution and different socioeconomic groups.

Several important conclusions can be drawn from this analysis. As already acknowledged in the literature, the ETI is not a structural parameter and this study shows that policy conclusions can be misleading. Reported estimates need to be interpreted within the context they are estimated in and researchers and policy makers need to be careful about what type and size of elasticity should be used for policy analysis (e.g. when calibrating an optimal tax model). In addition, the analysis shows that results are sensitive. It is remarkable how many details about the (constructed) income measures are missing in primary studies (e.g. what is included in the tax base). To develop new (empirical) strategies that are robust to certain model choices, we need to raise the awareness that insignificant and even implausible estimates are meaningful.

References

Aarbu, K. O. and Thoresen, T. O. (2001), 'Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform', *National Tax Journal* **54**(2), 319–335.

Alvaredo, F., Atkinson, A. B., Piketty, T. and Saez, E. (2013), 'The top 1 percent in international and historical perspective', *Journal of Economic perspectives* **27**(3), 3–20.

Arrazola, M., de Hevia, J., Romero, D. and Sanz-Sanz, J. F. (2014), 'Personal Income Tax Reforms and the Elasticity of Reported Income to Marginal Tax Rates: An Empirical Analysis Applied to Spain', *Working Paper Series in Public Finance provided by Victoria Business School* .

Arrazola-Vacas, M., Sanz-Sanz, J. F., , Rueda-Lopez, N. and Romero-Jordan, D. (2015), 'Reported Gross Income and Marginal Tax Rates: Estimation of the Behavioural Reactions of Spanish Taxpayers', *Applied Economics* **47**(5), 466–484.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999), 'A review of estimates of the schooling/earnings relationship, with tests for publication bias', *Labour economics* **6**(4), 453–470.

Auten, G. and Carroll, R. (1999), 'The Effect of Income Taxes on Household Income', *Review of Economics and Statistics* **81**(4), 681–693.

Auten, G., Carroll, R. and Gee, G. (2008), 'The 2001 and 2003 Tax Rate Reductions: An Overview and Estimate of the Taxable Income Response', *National Tax Journal* **61**(3), 345–364.

Auten, G. and Joulfaian, D. (2009), 'The Taxable Income Elasticity of High-Income Taxpayers: Evidence from a Long Panel', *Working Paper* .

Auten, G. and Kawano, L. (2014), 'How the Rich Respond to Anticipated Tax Increases: Evidence from the 1993 Tax Act', *Working Paper* .

Bakos, P., Benczúr, P. and Benedek, D. (2010), 'The elasticity of taxable income: Estimates and flat tax predictions using the hungarian tax changes in 2005', *Working Paper* .

Blomquist, S. and Selin, H. (2010), 'Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates', *Journal of Public Economics* **94**(11), 878–889.

Brodeur, A., Lé, M., Sangnier, M. and Zylberberg, Y. (2016), 'Star wars: The empirics strike back', *American Economic Journal: Applied Economics* **8**(1), 1–32.

Burlig, F. (2018), 'Improving transparency in observational social science research: A pre-analysis plan approach', *Economics Letters* **168**, 56 – 60.

Burns, S. K. and Ziliak, J. P. (2017), 'Identifying the elasticity of taxable income', *The Economic Journal* **127**, 297–329.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008), 'Bootstrap-based improvements for inference with clustered errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Card, D., Kluve, J. and Weber, A. (2010), 'Active labour market policy evaluations: A meta-analysis', *The Economic Journal* **120**(548), F452–F477.

Card, D., Kluve, J. and Weber, A. (2018), 'What works? a meta analysis of recent active labor market program evaluations', *Journal of the European Economic Association* **16**(3), 894–931.

Card, D. and Krueger, A. B. (1995), 'Time-series minimum-wage studies: a meta-analysis', *American Economic Review* **85**(2), 238–243.

Carey, S., Creedy, J., Gemmell, N. and Teng, J. (2015), 'Estimating the Elasticity of Taxable Income in New Zealand', *Economic Record* **91**(292), 54–78.

Carroll, R. (1998), 'Do Taxpayers Really Respond to Changes in Tax Rates? Evidence from the 1993 Tax Act', *Office of Tax Analysis Working Paper* (79), 145–77.

Carroll, R. and Hrung, W. (2005), 'What Does the Taxable Income Elasticity Say about Dynamic Responses to Tax Changes?', *American Economic Review* **95**(2), 426–431.

Chetty, R. (2009), 'Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance', *American Economic Journal: Economic Policy* **1**(2), 31–52.

Chetty, R. (2012), 'Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply', *Econometrica* **80**(3), 969–1018.

Chetty, R., Friedman, J. N., Olsen, T. and Pistaferri, L. (2011), 'Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records', *Quarterly Journal of Economics* **126**, 749–804.

Chetty, R., Guren, A., Manoli, D. S. and Weber, A. (2012), 'Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta- Analysis of Extensive Margin Elasticities', *NBER Macroeconomics Annual* **27**(1), 1–56.

Christensen, G. and Miguel, E. (2016), 'Transparency, Reproducibility, and the Credibility of Economics Research', *forthcoming, Journal of Economic Literature* .

Diamond, P. A. (1998), 'Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates', *American Economic Review* **88**(1), 83–95.

Diamond, P. A. and Saez, E. (2011), 'The Case for a Progressive Tax: From Basic Research to Policy Recommendations', *Journal of Economic Perspectives* **25**(4), 165–190.

Doerrenberg, P., Peichl, A. and Siegloch, S. (2017), 'The Elasticity of Taxable Income in the Presence of Deduction Possibilities', *Journal of Public Economics* **151**, 41–55.

Egger, M., Smith, G. D., Schneider, M. and Minder, C. (1997), 'Bias in meta-analysis detected by a simple, graphical test', *Bmj* **315**(7109), 629–634.

Ericson, P., Flood, L. and Islam, N. (2015), 'Taxes, Wages and Working Hours', *Empirical Economics* **49**(2), 503–535.

Fack, G. and Landais, C. (2016), 'The Effect of Tax Enforcement on Tax Elasticities: Evidence from Charitable Contributions in France', *Journal of Public Economics* **133**, 23–40.

Feld, L. P. and Heckemeyer, J. H. (2011), 'FDI and Taxation: A Meta-Study', *Journal of Economic Surveys* **25**(2), 233–272.

Feldstein, M. (1995), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act', *Journal of Political Economy* **103**(3), 551–572.

Feldstein, M. (1999), 'Tax Avoidance and the Deadweight Loss of the Income Tax', *Review of Economics and Statistics* **81**(4), 674–680.

Feldstein, M. and Feenberg, D. (1996), 'The Effect of Increased Tax Rates on Taxable Income and Economic Efficiency: A Preliminary Analysis of the 1993 Tax Rate Increases', *Tax Policy and the Economy* **10**, 89–117.

Fullerton, D. (1996), Comment on 'High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response', *in* 'Empirical Foundations of Household Taxation', pp. 189–92.

Gelber, A. M. (2014), 'Taxation and the Earnings of Husbands and Wives: Evidence from Sweden', *Review of Economics and Statistics* **96**(2), 287–305.

Giertz, S. H. (2007), 'The Elasticity of Taxable Income over the 1980s and 1990s', *National Tax Journal* pp. 743–768.

Giertz, S. H. (2010*a*), 'Panel Data Techniques and the Elasticity of Taxable Income', *National Tax Association 2009 Annual Conference Proceedings* .

Giertz, S. H. (2010*b*), 'The Elasticity of Taxable Income during the 1990s: New Estimates and Sensitivity Analyses', *Southern Economic Journal* **77**(2), 406–433.

Goolsbee, A. (2000), 'What Happens When You Tax the Rich? Evidence from Executive Compensation', *Journal of Political Economy* **108**(2), 352–378.

Gordon, R. H. and Kopczuk, W. (2014), 'The Choice of the Personal Income Tax Base', *Journal of Public Economics* **118**, 97–110.

Gordon, R. H. and Slemrod, J. (1998), 'Are "Real" Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?', *NBER Working Paper No. 6576* .

Gottfried, P. and Schellhorn, H. (2004), 'Empirical Evidence on the Effects of Marginal Tax Rates on Income–The German Case', *IAW-Diskussionspapiere* (15).

Gottfried, P. and Witczak, D. (2009), 'The Responses of Taxable Income Induced by Tax Cuts - Empirical Evidence from the German Taxpayer Panel', *IAW-Diskussionspapiere* (57).

Gruber, J. and Saez, E. (2000), 'The Elasticity of Taxable Income: Evidence and Implications', *NBER Working Paper No. 7512* .

Gruber, J. and Saez, E. (2002), 'The Elasticity of Taxable Income: Evidence and Implications', *Journal of Public Economics* **84**(1), 1–32.

Hansson, Å. (2007), 'Taxpayers' Responsiveness to Tax Rate Changes and Implications for the Cost of Taxation in Sweden', *International Tax and Public Finance* **14**(5), 563–582.

Hargaden, E. P. (2015), 'Taxpayer responses over the cycle: Evidence from irish notches', *Working Paper* .

Harju, J. and Matikka, T. (2016), 'The elasticity of taxable income and income-shifting: what is "real"and what is not?', *International Tax and Public Finance* **23**(4), 640–669.

Havránek, T. (2015), 'Measuring intertemporal substitution: The importance of method choices and selective reporting', *Journal of the European Economic Association* **13**(6), 1180–1204.

Heim, B. T. (2009), 'The Effect of Recent Tax Changes on Taxable Income', *Journal of Policy Analysis and Management* **28**(1), 147–163.

Heim, B. T. (2010), 'The Responsiveness of Self-Employment Income to Tax Rate Changes', *Labour Economics* **17**(6), 940–950.

Heim, B. T. and Mortenson, J. A. (2016), 'The effect of recent tax changes on taxable income: Correction and update', *Journal of Policy Analysis and Management* .

Holmlund, B. and Söderström, M. (2011), 'Estimating Dynamic Income Responses to Tax Reform', *The BE Journal of Economic Analysis & Policy* **11**(1), 1–38.

Jongen, E. L. and Stoel, M. (2013), 'Estimating the Elasticity of Taxable Labour Income in the Netherlands', *Working Paper* .

Kiss, A. and Mosberger, P. (2014), 'The Elasticity of Taxable Income of High Earners: Evidence from Hungary', *Empirical Economics* **48**(2), 883–908.

Kleven, H. J. (2014), 'How Can Scandinavians Tax So Much?', *Journal of Economic Perspectives* **28**(4), 77–98.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011), 'Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark', *Econometrica* **79**(3), 651–692.

Kleven, H. J., Kreiner, C. T. and Saez, E. (2016), 'Why can modern governments tax so much? an agency model of firms as fiscal intermediaries', *Economica* **83**(330), 219–246.

Kleven, H. J. and Schultz, E. A. (2014), 'Estimating Taxable Income Responses Using Danish Tax Reforms', *American Economic Journal: Economic Policy* **6**(4), 271–301.

Kopczuk, W. (2005), 'Tax Bases, Tax Rates and the Elasticity of Reported Income', *Journal of Public Economics* **89**(11), 2093–2119.

Kopczuk, W. (2015), 'Polish Business Flat Tax and its Effect on Reported Incomes: Preliminary Analysis', *Working Paper* .

Kreiner, C. T., Leth-Pedersen, S. and Skov, P. E. (2016), 'Tax Reforms and Intertemporal Shifting of Wage Income: Evidence from Danish Monthly Payroll Records', *American Economic Journal: Economic Policy, forthcoming* .

Kreiner, C. T., Leth-Petersen, S. and Skov, P. E. (2014), 'Year-End Tax Planning of Top Management: Evidence from High-Frequency Payroll Data', *American Economic Review* **104**(5), 154–158.

Kumar, A. and Liang, C.-Y. (2015), 'The Taxable Income Elasticity: A Structural Differencing Approach', *Working Paper* .

Le Maire, D. and Schjerning, B. (2013), 'Tax bunching, income shifting and self-employment', *Journal of Public Economics* **107**, 1–18.

Lehmann, E., Marical, F. and Rioux, L. (2013), 'Labor Income Responds Differently to Income-Tax and Payroll-Tax Reforms', *Journal of Public Economics* **99**, 66–84.

Lichter, A., Peichl, A. and Siegloch, S. (2015), 'The own-wage elasticity of labor demand: A meta-regression analysis', *European Economic Review* **80**(C), 94–119.

Lindsey, L. B. (1987), 'Individual Taxpayer Response to Tax Cuts: 1982–1984: With Implications for the Revenue Maximizing Tax Rate', *Journal of Public Economics* **33**(2), 173–206.

Looney, A. and Singhal, M. (2006), 'The Effect of Anticipated Tax Changes on Intertemporal Labor Supply and the Realization of Taxable Income', *Working Paper* .

Mankiw, N. G., Weinzierl, M. and Yagan, D. (2009), 'Optimal Taxation in Theory and Practice', *Journal of Economic Perspectives* **23**(4), 147–174.

Massarrat Mashhadi, N. and Werdt, C. (2012), 'Estimating Dynamic Income Responses to Tax Changes', *Working Paper* .

Matikka, T. (2016), 'The Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland', *Forthcoming, Scandinavian Journal of Economics.* .

Meghir, C. and Phillips, D. (2010), 'Labour supply and taxes', *Dimensions of tax design: The Mirrlees Review* pp. 202–74.

Mirrlees, J. A. (1971), 'An Exploration in the Theory of Optimum Income Taxation', *The Review of Economic Studies* **38**(2), 175–208.

Moffitt, R. and Wilhelm, M. O. (2000), 'Taxation and the labor supply decisions of the affluent', *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich* p. 193.

Mortenson, J. A. (2016), ‘All Income is Not Created Equal: Cross-Tax Elasticity Estimates in the United States’, *Working Paper* .

Mortenson, J. A. and Whitten, A. (2015), ‘How Sensitive Are Taxpayers to Marginal Tax Rates? Evidence from Income Bunching in the United States’, *Working Paper* .

Paetzold, J. (2017), ‘How do Wage Earners Respond to a Large Kink? Evidence on Earnings and Deduction Behavior from Austria.’, *Working Paper* .

Piketty, T. and Saez, E. (2013), ‘Optimal Labor Income Taxation’, *Handbook of Public Economics* **5**, 391–474.

Piketty, T., Saez, E. and Stantcheva, S. (2014), ‘Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities’, *American Economic Journal: Economic Policy* **6**(1), 1–44.

Pirttilä, J. and Selin, H. (2011*a*), ‘Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993’, *The Scandinavian Journal of Economics* **113**(1), 120–144.

Pirttilä, J. and Selin, H. (2011*b*), Tax Policy and Employment: How Does the Swedish System Fare? CESIFO Working Paper No.3355.

Romer, C. D. and Romer, D. H. (2010), ‘The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks’, *American Economic Review* **100**(3), 763–801.

Rosenthal, R. (1979), ‘The File Drawer Problem and Tolerance for Null Results.’, *Psychological Bulletin* **86**(3), 638.

Rothstein, H. R., Sutton, A. J. and Borenstein, M. (2006), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, John Wiley & Sons.

Saez, E. (2001), ‘Using Elasticities to Derive Optimal Income Tax Rates’, *Review of Economic Studies* **68**(1), 205–229.

Saez, E. (2003), ‘The Effect of Marginal Tax Rates on Income: a Panel Study of ’Bracket Creep’’, *Journal of Public Economics* **87**(5), 1231–1258.

Saez, E. (2010), 'Do Taxpayers Bunch at Kink Points?', *American Economic Journal: Economic Policy* **2**(3), 180–212.

Saez, E., Slemrod, J. B. and Giertz, S. H. (2009), 'The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review', *NBER Working Paper No. 15012* .

Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The Elasticity of Taxable Incomewith Respect to Marginal Tax Rates: A Critical Review', *Journal of Economic Literature* **50**(1), 3–50.

Saez, E. and Veall, M. R. (2005), 'The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence', *American Economic Review* **95**(3), 831–849.

Sammartino, F. and Weiner, D. (1997), 'Recent Evidence on Taxpayers' Response to the Rate Increases in the 1990s', *National Tax Journal* **50**(3), 683–705.

Schmidt, T.-P. and Müller, H. (2012), 'Die Elastizität des zu versteuernden Einkommens in Deutschland: Eine empirische Untersuchung auf Basis des deutschen Taxpayer-Panels', *Working Paper* .

Sillamaa, M.-A. and Veall, M. R. (2001), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1988 Tax Flattening in Canada', *Journal of Public Economics* **80**(3), 341–356.

Singleton, P. (2011), 'The Effect Of Taxes On Taxable Earnings: Evidence From The 2001 And Related U.S. Federal Tax Acts', *National Tax Journal* **64**(2), 323.

Slemrod, J. (1995), 'Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986', *American Economic Review* **85**(2), 175–180.

Slemrod, J. (1996), High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response, *in* 'Empirical Foundations of Household Taxation', University of Chicago Press, pp. 169–192.

Slemrod, J. (1998), 'Methodological Issues in Measuring and Interpreting Taxable Income Elasticities', *National Tax Journal* **51**(4), 773–788.

Slemrod, J. (2007), 'Cheating Ourselves: The Economics of Tax Evasion', *Journal of Economic Perspectives* **21**(1), 25–48.

Slemrod, J. (2016), 'Caveats to the Research Use of Tax-Return Administrative Data', *National Tax Journal* **69**(4), 1003.

Slemrod, J. and Kopczuk, W. (2002), 'The Optimal Elasticity of Taxable Income', *Journal of Public Economics* **84**(1), 91–112.

Stanley, T. D. and Doucouliagos, H. (2017), 'Neither fixed nor random: weighted least squares meta-regression', *Research Synthesis Methods* **8**(1), 19–42.

Stanley, T., Doucouliagos, C., Giles, M., Heckemeyer, J., Johnston, R., Laroche, P., Nelson, J., Paldam, M., Poot, J., Pugh, G. et al. (2013), 'Meta Analysis of Economics Research Reporting Guidelines', *Journal of Economic Surveys* **27**(2), 390–394.

Stanley, T. and Doucouliagos, H. (2010), 'Picture this: a simple graph that reveals much ado about research', *Journal of Economic Surveys* **24**(1), 170–191.

Sterne, J. A., Harbord, R. M. et al. (2004), 'Funnel plots in meta-analysis', *Stata Journal* **4**, 127–141.

Thomas, A. (2012), 'The Elasticity of Taxable Income in New Zealand: Evidence from the 1986 Tax Reform', *New Zealand Economic Papers* **46**(2), 159–167.

Thoresen, T. O. and Vattø, T. E. (2015), 'Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature', *Labour Economics* **37**, 38–53.

Weber, C. (2014*a*), 'Measuring Treatment for Tax Policy Analysis', *Working Paper* .

Weber, C. (2014*b*), 'Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences', *Journal of Public Economics* **117**, 90–103.

Werdt, C. (2015), 'The Elasticity of Taxable Income for Germany and its Sensitivity to the Appropriate Model', *Working Paper* .

Appendices

A Meta (Estimation) Sample

A.1 List of Included Studies

Aarbu, K. O. and Thoresen, T. O. (2001), 'Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform', *National Tax Journal* 54(2), 319–335.

Arrazola, M., de Hevia, J., Romero, D. and Sanz-Sanz, J. F. (2014), 'Personal Income Tax Reforms and the Elasticity of Reported Income to Marginal Tax Rates: An Empirical Analysis Applied to Spain', *Working Paper Series in Public Finance provided by Victoria Business School (September 2014)*.

Arrazola-Vacas, M., Sanz-Sanz, J. F., , Rueda-Lopez, N. and Romero-Jordan, D. (2015), 'Reported Gross Income and Marginal Tax Rates: Estimation of the Behavioural Reactions of Spanish Taxpayers', *Applied Economics* 47(5), 466–484.

Auten, G. and Carroll, R. (1999), 'The Effect of Income Taxes on Household Income', *Review of Economics and Statistics* 81(4), 681–693.

Auten, G., Carroll, R. and Gee, G. (2008), 'The 2001 and 2003 Tax Rate Reductions: An Overview and Estimate of the Taxable Income Response', *National Tax Journal* 61(3), 345–364.

Auten, G. and Joulfaian, D. (2009), 'The Taxable Income Elasticity of High-Income Taxpayers: Evidence from a Long Panel', *Working Paper (May 2009)*.

Auten, G. and Kawano, L. (2014), 'How the Rich Respond to Anticipated Tax Increases: Evidence from the 1993 Tax Act', *Working Paper (May 2014)*.

Bakos, P., Benczúr, P. and Benedek, D. (2010), 'The Elasticity of Taxable Income: Estimates and Flat Tax Predictions using the Hungarian Tax Changes in 2005', *Working Paper (September 2010)*.

Blomquist, S. and H. Selin (2010), 'Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates', *Journal of Public Economics* 94 (11-12), 878-889.

Burns, S. K. and Ziliak, J. P. (2017), 'Identifying the elasticity of taxable income', *The Economic Journal* 127, 297–329.

Carey, S., Creedy, J., Gemmell, N. and Teng, J. (2015), 'Estimating the Elasticity of Taxable Income in New Zealand', *Economic Record* 91(292), 54–78.

Carroll, R. (1998), 'Do Taxpayers Really Respond to Changes in Tax Rates? Evidence from the 1993 Tax Act', *Office of Tax Analysis Working Paper* (79), 145–77 (November 1998).

Chetty, R., Friedman, J. N., Olsen, T. and Pistaferri, L. (2011), 'Adjustment Costs, Firm

Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records', *Quarterly Journal of Economics* 126, 749–804.

Doerrenberg, P., Peichl, A. and Siegloch, S. (2017), 'The Elasticity of Taxable Income in the Presence of Deduction Possibilities', *Journal of Public Economics* 151, 41–55.

Ericson, P., Flood, L. and Islam, N. (2015), 'Taxes, Wages and Working Hours', *Empirical Economics* 49(2), 503–535.

Gelber, A. M. (2014), 'Taxation and the Earnings of Husbands and Wives: Evidence from Sweden', *Review of Economics and Statistics* 96(2), 287–305.

Giertz, S. H. (2007), 'The Elasticity of Taxable Income over the 1980s and 1990s', *National Tax Journal* pp. 743–768.

Giertz, S.H. (2010), 'Panel Data Techniques and the Elasticity of Taxable Income', *National Tax Association 2009 Annual Conference Proceedings*

Giertz, S. H. (2010), 'The Elasticity of Taxable Income during the 1990s: New Estimates and Sensitivity Analyses', *Southern Economic Journal* 77(2), 406–433.

Gottfried, P. and Schellhorn, H. (2004), 'Empirical Evidence on the Effects of Marginal Tax Rates on Income – The German Case', *IAW-Diskussionspapiere* (15) (February 2004).

Gottfried, P. and Witczak, D. (2009), 'The Responses of Taxable Income Induced by Tax Cuts - Empirical Evidence from the German Taxpayer Panel', *IAW-Diskussionspapiere* (57) (November 2009).

Gruber, J. and Saez, E. (2002), 'The Elasticity of Taxable Income: Evidence and Implications', *Journal of Public Economics*, vol. 84(1), pp. 1–32.

Hansson, Å. (2007), 'Taxpayers' Responsiveness to Tax Rate Changes and Implications for the Cost of Taxation in Sweden', *International Tax and Public Finance* 14(5), 563–582.

Harju, J. and Matikka, T. (2016), 'The Elasticity of Taxable Income and Income-Shifting: What is "real"and what is not?', *International Tax and Public Finance* 23(4), 640–669.

Heim, B. T. (2010), 'The Responsiveness of Self-Employment Income to Tax Rate Changes', *Labour Economics* 17(6), 940–950.

Heim, B. T. and Mortenson, J. A. (2016), 'The Effect of Recent Tax Changes on Taxable Income: Correction and Update', *Journal of Policy Analysis and Management*.

Holmlund, B. and Söderström, M. (2011), 'Estimating Dynamic Income Responses to Tax Reform', *The BE Journal of Economic Analysis & Policy* 11(1), 1–38.

Jongen, E. L. and Stoel, M. (2013), 'Estimating the Elasticity of Taxable Labour Income in the Netherlands', *Working Paper (May 2013)*.

Kiss, A. and Mosberger, P. (2014), 'The Elasticity of Taxable Income of High Earners: Evidence from Hungary', *Empirical Economics* 48(2), 883–908.

Kleven, H. J. and Schultz, E. A. (2014), 'Estimating Taxable Income Responses Using Danish Tax Reforms', *American Economic Journal: Economic Policy* 6(4), 271–301.

Kopczuk, W. (2005), 'Tax Bases, Tax Rates and the Elasticity of Reported Income', *Journal of Public Economics* 89(11), 2093–2119.

Kopczuk, W. (2015), 'Polish Business Flat Tax and its Effect on Reported Incomes: Preliminary Analysis', *Working Paper (June 2015)*.

Kumar, A. and Liang, C.-Y. (2015), 'The Taxable Income Elasticity: A Structural Differencing Approach', *Working Paper (January 2015)*.

Lehmann, E., Marical, F. and Rioux, L. (2013), 'Labor Income Responds Differently to Income-Tax and Payroll-Tax Reforms', *Journal of Public Economics* 99, 66–84.

Lindsey, L. B. (1987), 'Individual Taxpayer Response to Tax Cuts: 1982–1984: With Implications for the Revenue Maximizing Tax Rate', *Journal of Public Economics* 33(2), 173–206.

Looney, A. and Singhal, M. (2006), 'The Effect of Anticipated Tax Changes on Intertemporal Labor Supply and the Realization of Taxable Income', *Working Paper (July 2006)*.

Massarrat Mashhadi, N. and Werdt, C. (2012), 'Estimating Dynamic Income Responses to Tax Changes', *Working Paper (December 2012)*.

Matikka, T. (2016), 'The Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland', Forthcoming, *Scandinavian Journal of Economics*.

Moffitt, R. and Wilhelm, M. O. (2000), 'Taxation and the Labor Supply Decisions of the Affluent', *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich* p. 193.

Mortenson, J. A. (2016), 'All Income is Not Created Equal: Cross-Tax Elasticity Estimates in the United States', *Working Paper (March 2016)*.

Pirttilä, J. and Selin, H. (2011a), 'Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993', *The Scandinavian Journal of Economics* 113 (1), 120–144.

Saez, E. (2003), 'The Effect of Marginal Tax Rates on Income: a Panel Study of 'Bracket Creep', *Journal of Public Economics* 87(5), 1231–1258.

Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review', *Journal of Economic Literature* 50(1), 3–50.

Schmidt, T.-P. and Müller, H. (2012), 'Die Elastizität des zu versteuernden Einkommens in Deutschland: Eine empirische Untersuchung auf Basis des deutschen Taxpayer-Panels', *Working Paper (June 2012)*.

Sillamaa, M.-A. and Veall, M. R. (2001), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1988 Tax Flattening in Canada', *Journal of Public Economics* 80(3), 341–356.

Singleton, P. (2011), 'The Effect Of Taxes On Taxable Earnings: Evidence From The 2001 And Related U.S. Federal Tax Acts', *National Tax Journal* 64(2), 323.

Thomas, A. (2012), 'The Elasticity of Taxable Income in New Zealand: Evidence from the 1986 Tax Reform', *New Zealand Economic Papers* 46(2), 159–167.

Thoresen, T. O. and Vattø, T. E. (2015), 'Validation of the Discrete Choice Labor Supply Model by Methods of the New Tax Responsiveness Literature', *Labour Economics* 37, 38–53

Weber, C. (2014a), 'Measuring Treatment for Tax Policy Analysis', *Working Paper (November 2014a)*.

Weber, C. (2014b), 'Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences', *Journal of Public Economics* 117, 90–103.

Werdt, C. (2015), 'The Elasticity of Taxable Income for Germany and its Sensitivity to the Appropriate Model', *Working Paper (Jan 2015)*.

A.2 Distribution of Estimates by Study:Published vs Working Paper

Published articles **Study**	**# estimates**	**in % of full sample**
Aarbu and Thoresen (2001)	8	0.55
Arrazola-Vacas et al. (2015) CHANGE	27	1.86
Auten and Carroll (1999)	20	1.38
Auten et al. (2008)	10	0.69
Blomquist and Selin (2010)	10	0.69
Burns and Ziliak (2017)	68	4.70
Carey et al. (2015)	6	0.41
Chetty et al. (2011)	6	0.41
Doerrenberg et al. (2017)	16	1.10
Ericson et al. (2015)	5	0.35
Gelber (2014)	16	1.10
Giertz (2007)	69	4.77
Giertz (2010*b*)	127	8.77
Gruber and Saez (2002)	35	2.42
Hansson (2007)	30	2.07
Harju and Matikka (2016)	14	0.97
Heim (2010)	14	0.97
Heim and Mortenson (2016)	14	0.97
Holmlund and Söderström (2011)	36	2.49
Kiss and Mosberger (2014)	15	1.04
Kleven and Schultz (2014)	114	7.87
Kopczuk (2005)	91	6.28
Lehmann et al. (2013)	18	1.24
Lindsey (1987)	14	0.97
Matikka (2016)	18	1.24
Moffitt and Wilhelm (2000)	39	2.69
Pirttilä and Selin (2011*b*)	10	0.69
Saez (2003)	91	6.28
Saez et al. (2012)	24	1.66
Sillamaa and Veall (2001)	25	1.73
Singleton (2011)	25	1.73
Thomas (2012)	8	0.55
Thoresen and Vattø (2015)	21	1.45
Weber (2014*b*)	35	2.42
Total (published)	1061	73.27 %
Working Papers **Study**	**# estimates**	**in % of full sample**
Arrazola et al. (2014)	8	0.55
Auten and Joulfaian (2009)	24	1.66
Auten and Kawano (2014)	12	0.83
Bakos et al. (2010)	21	1.45
Carroll (1998)	12	0.83
Giertz (2010*a*)	72	4.97
Gottfried and Schellhorn (2004)	11	0.76
Gottfried and Witczak (2009)	15	1.04
Jongen and Stoel (2013)	54	3.73
Kopczuk (2015)	30	2.07
Kumar and Liang (2015)	10	0.69
Looney and Singhal (2006)	15	1.04
Massarrat Mashhadi and Werdt (2012)	9	0.62
Mortenson (2016)	42	2.90
Schmidt and Müller (2012)	18	1.24
Weber (2014*a*)	5	0.35
Werdt (2015)	11	0.76
Total (Working paper)	387	26.73 %
Total (all estimates)	1448	100 %

Note: The data covers only observations with a given or calculable standard error. # estimates denote the number of estimates collected in a particular study and the corresponding percentage share shows the share a study has in the final sample.

B Additional Descriptives

B.1 Summary Statistics by Income Concept

Table 8: Distributions of Estimates by Income Concept

Tax Base	Mean	Median	Std. Dev.	Obs.	Studies
Before Deductions	0.296	0.203	1.269	852	38
Adjusted Gross Income	0.410	0.475	2.573	179	
Gross Income	0.315	0.236	0.556	371	
Earned Income	0.125	0.062	0.257	129	
Self employed Income	0.675	0.858	0.510	20	
Wage Income	0.211	0.138	0.602	153	
After Deductions	0.390	0.343	0.583	596	37
Taxable Income	0.384	0.330	0.603	553	
Taxable Earnings	0.445	0.444	0.186	43	
Total	0.34	0.260	1.044	1448	53

Note: The data covers only observations with a given or calculable standard error.

B.2 Distribution of Estimates by Country and Income Concepts

Table 9: Income Concepts by Country

Variable	Adj. G. Income	Gross Income	Taxable Income	Earned Income	Self employed	Wage Income	Taxable Earnings	Total
Canada	15	2	2	2	2	2	0	25
Denmark	0	18	18	78	0	6	0	120
Finland	0	6	17	0	0	19	0	42
France	0	0	0	0	0	18	0	18
Germany	3	20	57	0	0	0	0	80
Hungary	0	0	36	0	0	0	0	36
Netherlands	0	0	0	0	0	54	0	54
New Zealand	0	0	10	0	0	4	0	14
Norway	0	0	8	21	0	0	0	29
Poland	0	30	0	0	0	0	0	30
Spain	0	35	0	0	0	0	0	35
Sweden	12	26	17	12	0	0	30	97
USA	149	234	388	16	18	50	13	868
Total	179	371	553	129	20	153	43	1,448

Note: The sample covers only observations with a given or calculable standard error.

B.3 Distribution of Estimates by Year of Publication

Table 10: Year of Publication and Published Type

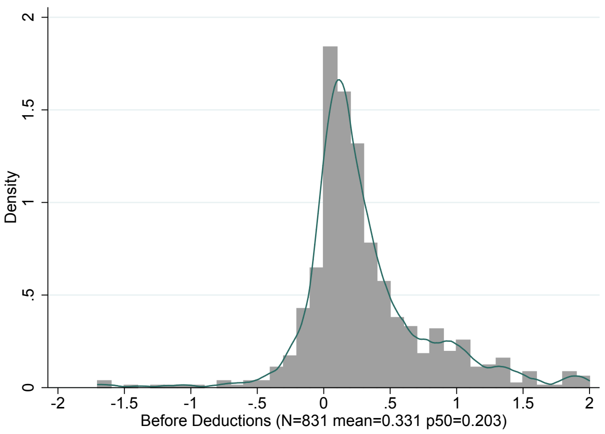
Year of Publication	Working Paper	Published	Total
1987	0	14	14
1998	12	0	12
1999	0	20	20
2000	0	39	39
2001	0	33	33
2002	0	35	35
2003	0	91	91
2004	11	0	11
2005	0	91	91
2006	15	0	15
2007	0	99	99
2008	72	10	82
2009	39	0	39
2010	21	151	171
2011	0	77	77
2012	27	32	59
2013	54	18	72
2014	25	191	216
2015	69	78	147
2016	42	82	124
Total	387	1061	1448

Note: The sample covers only observations with a given or calculable standard error.

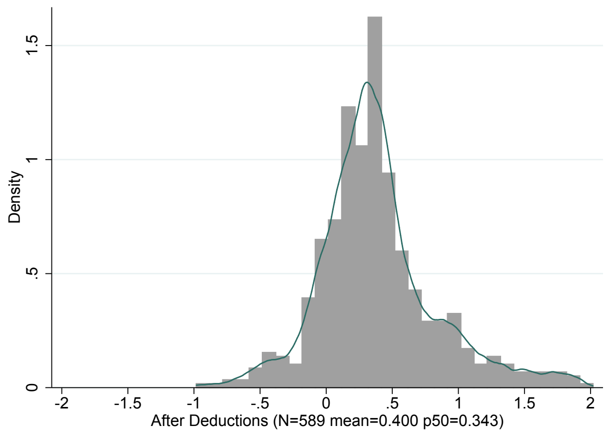
C Distribution of Elasticities and Details on Explanatory Variables

C.1 Distribution of Elasticities

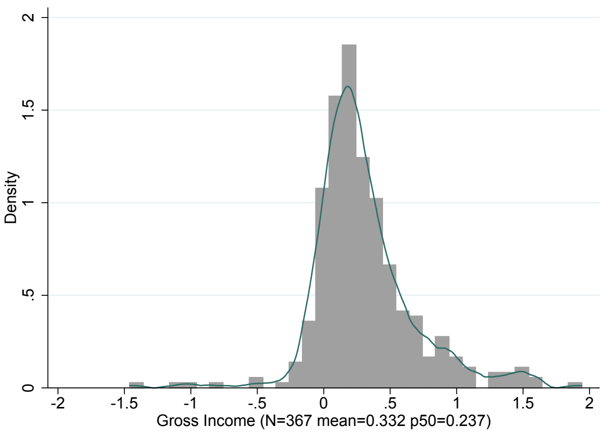
Figure 4: Distribution of Elasticities



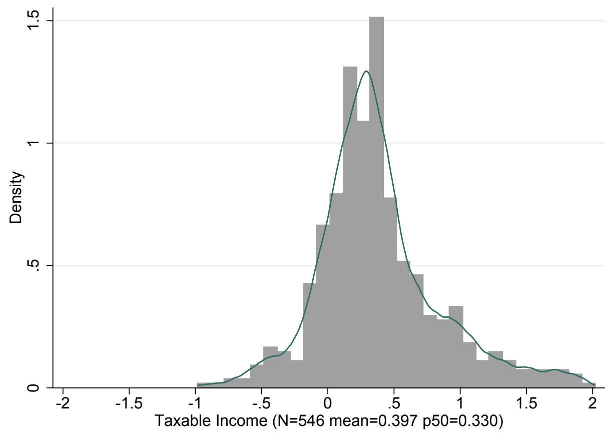
(a) Before Deductions



(b) After Deductions



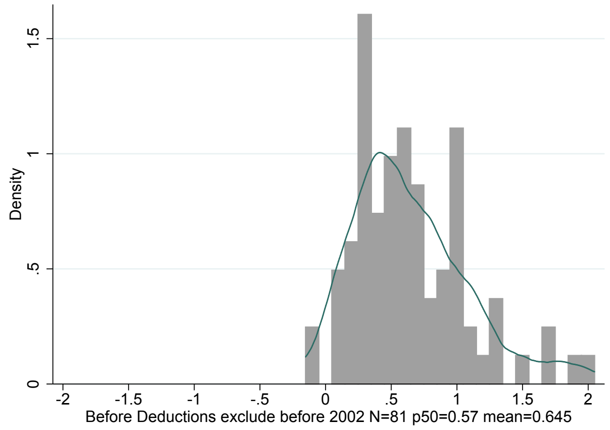
(c) Gross Income Elasticities



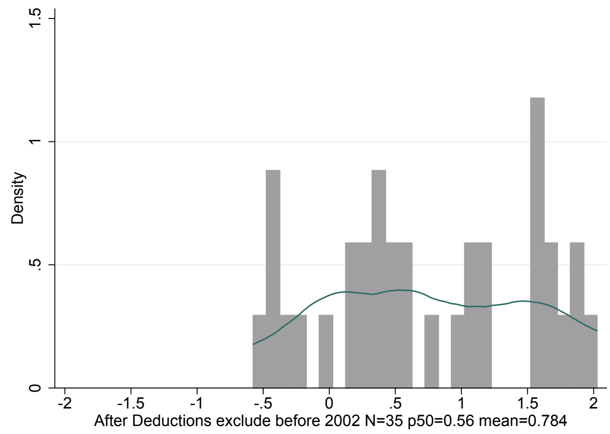
(d) Taxable Income Elasticities

Note: The data cover only observations with a given standard error or z-statistic. I restrict the sample to elasticity estimates that belong to the (a) before deductions subsample or (b) the after deduction subsample. Subfigures (c) and (d) are based on a narrower definition (gross or taxable income respectively).

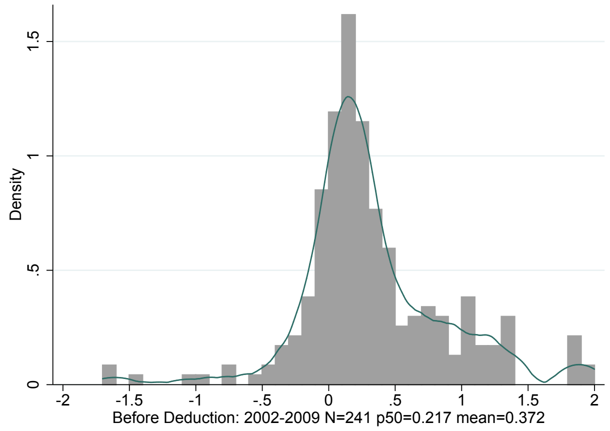
Figure 5: Distribution of Estimates by Publication Decade.



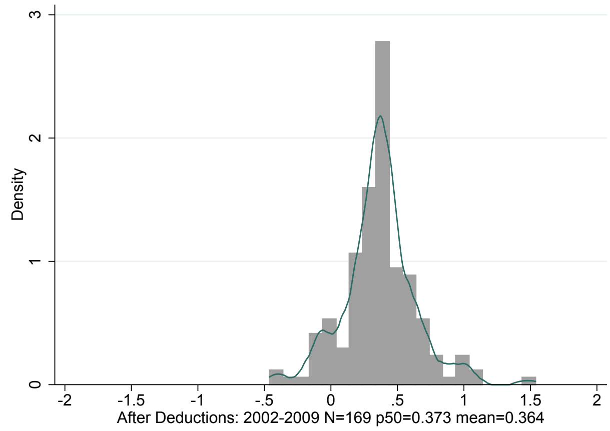
(a) Before Deductions <2002



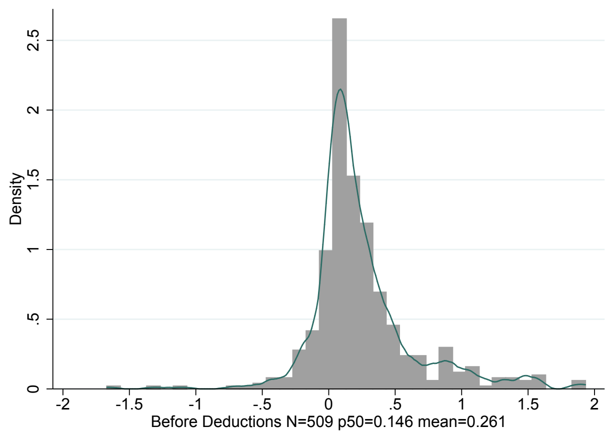
(b) After Deductions <2002



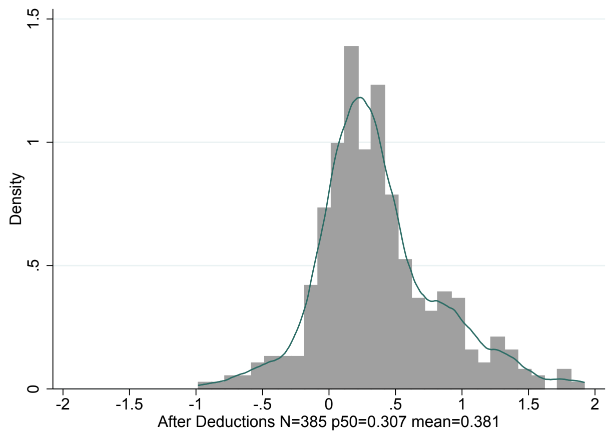
(c) Before Deductions >=2002 and <2009



(d) After Deductions >=2002 and <2009



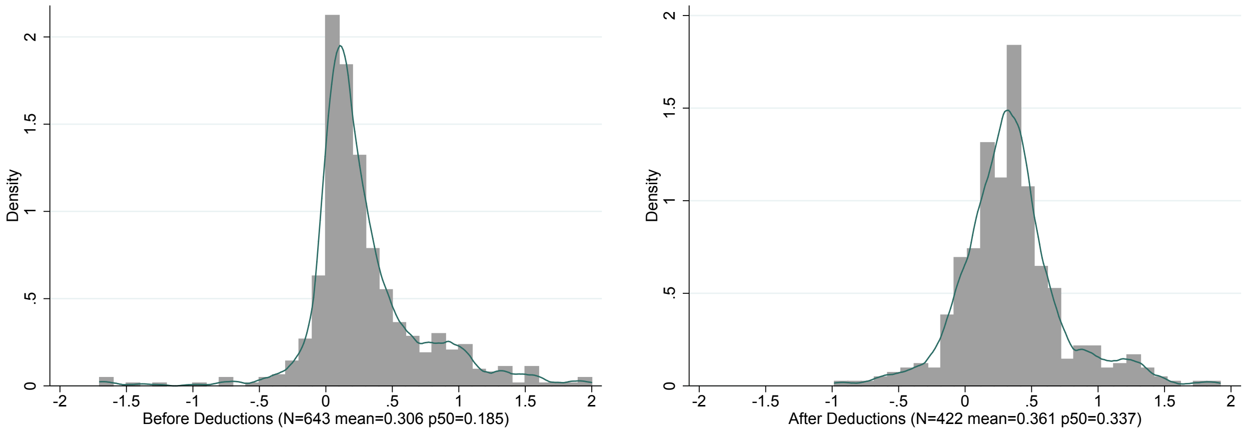
(e) Before Deductions >=2009



(f) After Deductions >=2009

Note: All graphs plot the distribution of elasticities by subsample and publication decade.

Figure 6: Distribution of Estimates (only income control(s)).



(a) Before Deductions (b) After Deductions

Note: Both graphs plot the distribution of elasticities that are derived with a specification using income control(s).

C.2 Explanatory Variables: Details

Regression technique: Most approaches use an Instrument for Δ NTR = $ln\left[\frac{(1-\tau_{it}(z_{it}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$ to achieve a causal relationship:

IV: mechanical tax rate changes: $\Delta ln(1-\tau_{it}^{p}) = ln\left[\frac{(1-\tau_{it}^{p}(z_{it-k}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$, where τ_{it}^{p} is the marginal tax rate that an individual would face given her synthetic income. Example: In year 3, τ_{it}^{p} would be calculated based on income of year two (assume time length of one year). Introduced by Auten and Carroll (1999) / Gruber and Saez (2002) and often referred to as the most standard specification.

IV: (lagged) mechanical tax rate changes: $\Delta ln(1-\tau_{it}^{p,lag})$, where $\tau_{it}^{p,lag}$ is based on income further in the past. $\Delta ln(1-\tau_{it}^{p}, lag) = ln\left[\frac{(1-\tau_{it}^{p}(z_{it-k-lag}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$

IV: other: This category summarizes all other instruments. (1) Blomquist and Selin (2010): They use a single difference and an imputed taxable income $\hat{z_{it}}$ to calculate their instrument: $\left(\frac{1-\tau_{it}(\hat{z_{it}})}{1-\tau_{it-k}(\hat{z_{it-k}})}\right)$. (2) Burns and Ziliak (2017): use a grouping estimator/ instrument. (3) Carey et al. (2015): Two instruments based on a time period with no tax changes to estimate dynamics of taxable income. (4) Carroll (1998): proxy for permanent income and calculate synthetic tax rate. (5) Ericson et al. (2015): instrument based on individual/ household-specific variables/ no measure of previous or future taxable income. (6) Harju and Matikka (2016): use Gruber and Saez (2002) and Weber (2014*b*) but include separate NTR for wage and dividend (plus, separate instruments). (7) Holmlund and Söderström (2011): use a dynamic model to explicitly measure short and long run responses. (8) Looney and Singhal (2006): NTR change based on family income stays the same; predict the change in marginal tax rates faced by families assuming that family income remains constant in real terms between year 1 and year 2. (9) Matikka (2016): use changes in flat municipal income tax rates as an instrument for overall changes in marginal tax rates. This instrument is not a function of individual income, which is the basis for an exogenous instrument. (10) Gelber (2014) explicitly control for NTR for wife and husband and extend the most standard specification to allow each spouse's earnings to depend not only on his or her own tax rate and unearned income, but also on the tax rate and unearned income of the other spouse.

DID and IV: Combination of a classical DID and an IV- estimation procedure. The instrument is a binary dummy variable. It determines treatment and control. (e.g. Saez, 2003 or Kopczuk, 2015) **DID classic**.

Income Controls: For the majority of coded specifications, there is no information available about what type of income (e.g. gross or taxable) is used.
Auten and Carroll (1999): '**Auten Carroll**' describes the use of log base year income $ln(z_{i,t-k})$ as an income control.
Mostly old studies and robustness checks deliver estimates that use no income control (**none)** at all.
Gruber and Saez (2002): '**Gruber Saez**' defines the inclusion of a spline of base year income as an income control.
Kopczuk (2005): '**Kopczuk**' defines the inclusion of two income control variables. The deviation of log base year income and lagged base year income and lagged base year income separately. To be more precise: $ln(z_{i,t-k-1})$, $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, spline of $ln(z_{i,t-k-1})$, spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of $ln(z_{i,t-k-1})$ and $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of $ln(z_{i,t-k})$ and spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of spline of $ln(z_{i,t-k})$ and $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$ and combination of spline of $ln(z_{i,t-k})$ and spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$.
The category '**other**' involves all other kinds of income controls. Example: Burns and Ziliak (2017) use a cohort-state-year income control in some specifications.

Difference Length The term difference length defines the time window k. If researchers relate 2005 to 2002, the time window will be 3 years.

Weighting by Income: This is a dummy variable that indicates whether (primary) estimation results are weighted by income.

Sample Restrictions:
Age Cutoff: It is a dummy variable that indicates whether an age cutoff is used.
Income Cutoff: I create subcategories: 0-10k, 10-12k, 12-31k and none. Some researchers do not apply any kind of income restrictions. However, sometimes it is not clear if they simply do not mention them, applied no income restriction on purpose or if their dataset considers a subgroup of tax-units in the first place. It often remains unclear what type of income is used (e.g. taxable or gross) to restrict the sample. I coded the values in national currency and recalculated them in US-Dollar. Purchasing power parities do not lead to different results.
Employment type: I distinguish between no restriction with respect to employment type (none), only wage earner, and only self employed individuals.
Marital Status: I distinguish between no restriction with respect to marital status (none), only married tax-units and only singles.

Variations across time and country:
Country Group: USA, Scandinavia (Denmark, Norway, Sweden) and Rest (Canada, Finland, France, Germany, Hungary, Netherlands, New Zealand, Poland, Spain)
Mean year in study data: I calculate the (rounded) mean year of observation based on time start and time end of dataset.
Estimation/ Data Decade: I used the mean year of the study data and assigned the respective decade: $<$ 1990, 1990-2000 and $>=$ 2000.

Publication Characteristics:
Publication Decade: 2001-1010, $<=$ 2000 and $>$ 2011.
Published Type: I distinguish between (1) published in a peer reviewed journal and (2) Working Paper.

Extension: Contextual Variables: For a particular estimate, I compare start and end year of (restricted) data period and add the tax related characteristics. Economy related characteristics are merged via the mean year of observation.
Tax Reform Characteristics: It is difficult and almost impossible to code precisely if taxes are increased, and if so, by how much. As an example, think of an estimate that uses data from 2001 to 2010 and exploits three tax changes at different points in the income distribution which differ additionally in magnitude. Therefore, I decided to focus only on two aspects: (1) introduction of a top tax bracket and (2) a reduction of tax brackets.
Intro of top tax bracket: information if reform involves an introduction of top. Source: Paper itself plus OECD Tax Database
Reduce number of tax brackets: information if reform involves a reduction of tax brackets. There are tax systems (e.g. the German tax system) that do not have a bracket system; these systems automatically receive a zero. Source: Paper itself plus OECD Tax Database
Economy related characteristics merged via link to mean year of observation (= use start and end year of (restricted) data period for collected primary estimate:
Gini (disposable income, post taxes and transfers) / Income Definition till 2011. The data is available since 1987. In rare cases I gave an observation the 1987-gini coefficient. To improve (regression) interpretation, I standardized the Gini Coefficient by multiplying it with 100. Source: http://stats.oecd.org (07.11.2016)
Top Income Shares: Pre-tax national income share held by a given percentile group (here top 1% and top 10%). Pre-tax national income is the sum of all pre-tax personal income flows accruing to the owners of the production factors, labour and capital, before taking into account the operation of the tax/transfer system, but after taking into account the operation of pension system. Source: World Inequality Database (extracted 16.07.2018)
Unemployment Rate: The unemployment rate is the number of unemployed people as a percentage of the labour force, with the latter consisting of the unemployed plus those in paid or self-employment. Unemployed people are those who report that they are out of work, that they are available for work and that they have taken active steps to find work in the last four weeks. When unemployment is high, some people become discouraged and stop looking for work; they are then excluded from the labour force. This implies that the unemployment rate may fall, or stop rising, even though there has been no underlying improvement in the labour market. (Source: OECD, Short-Term Labour Market Statistics; extracted 17.07.2018.)
Fraction self-employed: fraction self-employed is defined crudely as all non employees (self-employed, employers, and non classifiable workers) as a fraction of the workforce. Source: Kleven (2014)
Modern taxes / GDP: Kleven et al. (2016) decompose the tax take (=tax/GDP) into modern and traditional taxes. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value added taxes. Traditional taxes include all the other taxes. Source: Kleven et al. (2016)

Table 11: Descriptive Statistics: Sources of Heterogeneity

	Before Deductions (BD) (N=852) # studies	**After Deductions (AD) (N=596)** # studies
Estimation Techniques		
Regression technique		
IV: mechanical tax rate changes	29	25
IV: (lagged) mechanical tax rate changes	7	7
IV: other	6	8
DID and IV	7	3
classic DID	1	3
Income Control		
Auten Carroll (1999)	23	18
none	28	23
Gruber Saez (2002) spline	18	10
Kopczuk (2005) type	14	14
other	7	3
Difference Length		
3 years	17	20
1 year	23	20
2 years	11	8
4+ years	9	7
Weighted by Income	13	11
Sample Restrictions		
Age Cutoff	20	22
Income Cutoff		
0-10k	14	11
none	8	9
10k-12k	15	9
12-31k	17	12
$>$ 31k	21	18
Variations across Countries and Time		
Country Group		
USA	20	19
Scandinavia	5	6
other countries	13	12
Mean year in study data		
Estimation decade		
$<$ 1999	15	16
1990 - 2000	15	10
$>$ 2000	13	15
Publication Characteristics		
Publication decade		
2001-2010	11	15
$<=$ 2000	3	3
$>$ 2011	24	19
Published Type		
published in peer reviewed journal	25	23
working paper	13	14

Note: see text for description of sample. I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic. Reference categories are given in italics.

D Additional Sample Restrictions - Before Deductions (BD) and After Deductions

Researchers often conduct subgroup analysis by *marital status* or *employment type*. Single taxpayers might respond differently than married couples and it is obvious that a self-employed person has more control over his or her income compared to someone receiving only wage income.

Table 12: Descriptive Statistics: Sample Restrictions

	Before Deductions (BD) (N=832)		**After Deductions (AD) (N=589)**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Sample Restrictions				
Employment type				
none	0.743	0.437	0.934	0.249
wage earner	0.184	0.388	0.039	0.194
self-employed	0.073	0.261	0.027	0.163
Marital Status				
none	0.839	0.368	0.854	0.353
married	0.117	0.321	0.102	0.303
single	0.044	0.206	0.044	0.206

Note: see text for description of sample. I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic.

In line with expectations, a BD elasticity estimated on a subsample of only wage earners leads to a lower elasticity compared to a specification with no restriction on employment type. Greater coverage of third party information reporting and the associated lower evasion opportunities might be a reason (Kleven et al., 2011). If primary studies restrict their sample according to marital status, it appears that single taxpayers reveal a lower BD elasticity compared to no restriction.

Table 13: WLS before deductions results with add. sample restrictions

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.050*	0.035	0.051*	0.034	0.008	0.013
	(0.025)	(0.024)	(0.026)	(0.024)	(0.014)	(0.008)
IV-other	0.067	0.073	0.062	0.065	0.046	0.050
	(0.063)	(0.056)	(0.061)	(0.054)	(0.057)	(0.052)
DID-IV	0.289***	0.291***	0.283***	0.286***	0.291***	0.291***
	(0.061)	(0.083)	(0.064)	(0.087)	(0.058)	(0.070)
DID-classic	0.321***	0.159**	0.302***	0.147	0.164**	0.132
	(0.070)	(0.076)	(0.084)	(0.089)	(0.065)	(0.081)
Income Control (omitted: Auten Carroll)						
none	-0.210***	-0.207***	-0.209***	-0.207***	-0.204***	-0.204***
	(0.027)	(0.029)	(0.027)	(0.029)	(0.031)	(0.031)
Gruber Saez Spline	-0.018***	-0.012*	-0.018***	-0.013**	-0.009	-0.010
	(0.005)	(0.006)	(0.005)	(0.006)	(0.008)	(0.007)
Kopczuk	-0.016**	-0.007	-0.017**	-0.007	-0.006	-0.006
	(0.007)	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)
other	-0.034*	-0.008	-0.032	-0.003	-0.003	0.003
	(0.020)	(0.010)	(0.020)	(0.011)	(0.008)	(0.009)
Difference Length (omitted: 3-years)						
1 year	0.069	0.041	0.067	0.037	0.034	0.028
	(0.074)	(0.058)	(0.072)	(0.057)	(0.052)	(0.051)
2 years	-0.010	-0.043***	-0.012	-0.045***	-0.057***	-0.060***
	(0.024)	(0.014)	(0.025)	(0.015)	(0.015)	(0.022)
4 years and more	0.079	-0.003	0.081	-0.003	-0.017	-0.019
	(0.054)	(0.016)	(0.055)	(0.016)	(0.022)	(0.026)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied		-0.190***		-0.192***		-0.077
		(0.051)		(0.051)		(0.073)
Income Cutoff applied (omitted: 0-10k)						
none		-0.025		-0.025		-0.005
		(0.020)		(0.022)		(0.018)
10k-12k		-0.010		-0.010		-0.013
		(0.011)		(0.011)		(0.011)
12k-31k		0.010		0.012		0.008
		(0.007)		(0.010)		(0.012)
>31k		0.009		0.008		0.001
		(0.020)		(0.021)		(0.016)
Employment Type (omitted: no restriction)						
wage earner			-0.008*	-0.011**		-0.009**
			(0.005)	(0.005)		(0.004)
self-employed			-0.001	-0.003		0.000
			(0.004)	(0.004)		(0.004)
Marital Status (omitted: no restriction)						
married			0.017	0.010		0.013
			(0.029)	(0.033)		(0.035)
sinlge			0.005	-0.022**		-0.028***
			(0.023)	(0.009)		(0.008)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia					-0.056	0.020
					(0.045)	(0.080)
other countries					0.146**	0.163**
					(0.060)	(0.072)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.247*	0.295*
					(0.140)	(0.158)
after 2010					-0.143*	-0.163**
					(0.073)	(0.066)
Constant	0.072***	0.252***	0.077***	0.260***	0.258***	0.285***
	(0.006)	(0.055)	(0.009)	(0.054)	(0.050)	(0.053)
Observations	852	852	852	852	852	852
Adjusted R^2	0.573	0.636	0.573	0.637	0.660	0.661

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: WLS after deductions results with add. sample restrictions

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.447***	0.480**	0.449***	0.480**	0.332***	0.348**
	(0.129)	(0.224)	(0.131)	(0.225)	(0.108)	(0.146)
IV-other	-0.358***	-0.314***	-0.293***	-0.379***	0.255	-0.235
	(0.121)	(0.096)	(0.102)	(0.110)	(0.178)	(0.175)
DID-IV	-0.677***	-0.954***	-0.556***	-0.796***	-0.354	-0.454*
	(0.178)	(0.228)	(0.149)	(0.222)	(0.295)	(0.258)
DID-classic	0.129	0.051	0.132	0.048	0.179*	0.200***
	(0.114)	(0.105)	(0.112)	(0.105)	(0.090)	(0.071)
Income Control (omitted: Auten Carroll)						
none	0.128	-0.041	0.128	-0.043	0.008	-0.179*
	(0.126)	(0.100)	(0.128)	(0.102)	(0.147)	(0.091)
Gruber Saez Spline	-0.033	-0.093	-0.032	-0.095	-0.094	-0.189*
	(0.105)	(0.107)	(0.103)	(0.109)	(0.138)	(0.109)
Kopczuk-type	-0.380***	-0.471***	-0.391***	-0.468***	0.111	-0.152
	(0.114)	(0.107)	(0.131)	(0.107)	(0.116)	(0.140)
other	-0.327*	-0.568***	-0.297	-0.578***	0.105	-0.374**
	(0.180)	(0.147)	(0.198)	(0.164)	(0.144)	(0.180)
Difference Length (omitted: 3-years)						
1 year	0.026	0.193***	0.030	0.193***	0.007	0.155**
	(0.133)	(0.045)	(0.138)	(0.045)	(0.155)	(0.064)
2 years	0.199**	0.088	0.203**	0.072	0.229**	-0.015
	(0.077)	(0.094)	(0.077)	(0.098)	(0.091)	(0.088)
4 years and more	0.258	-0.027	0.262	-0.020	-0.302	-0.311
	(0.171)	(0.200)	(0.170)	(0.202)	(0.196)	(0.193)
Sample restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied		0.504***		0.512***		0.484***
		(0.077)		(0.079)		(0.118)
Income Cutoff applied (omitted: 0-10k)						
none		0.359***		0.373***		0.462***
		(0.062)		(0.072)		(0.057)
10k-12k		0.255**		0.263**		0.647**
		(0.114)		(0.117)		(0.243)
12k-31k		0.141**		0.147**		0.040
		(0.066)		(0.069)		(0.078)
>31k		1.061***		1.073***		1.091***
		(0.262)		(0.257)		(0.233)
Employment Type (omitted: no restriction)						
wage earner			0.029	-0.009		-0.010
			(0.055)	(0.016)		(0.012)
self-employed			-0.165	-0.357		-0.362
			(0.156)	(0.315)		(0.305)
Marital Status (omitted: no restriction)						
married			-0.091	0.033		0.247
			(0.121)	(0.127)		(0.173)
single			-0.029	0.105		0.330*
			(0.125)	(0.132)		(0.176)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia					0.025	0.239
					(0.098)	(0.241)
other countries					0.227**	0.561**
					(0.091)	(0.217)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.717*	0.931
					(0.413)	(0.560)
after 2010					-0.439***	-0.153
					(0.141)	(0.141)
Constant	0.443***	0.030	0.441***	0.025	0.359**	-0.355
	(0.113)	(0.091)	(0.110)	(0.089)	(0.134)	(0.252)
Observations	596	596	596	596	596	596
Adjusted R^2	0.580	0.729	0.579	0.731	0.642	0.778

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Contextual Factors - Full Results

E.1 Contextual Factors - Before Deductions (BD) - Full Results

Table 15: WLS before deductions - Contextual Variables

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reg. Technique (omitted: IV: Δ mechanical tax rate)								
IV: lagged Δ mechanical tax rate	0.050*	0.006	0.055	0.046**	0.031***	0.056*	0.043**	0.036**
	(0.025)	(0.015)	(0.037)	(0.022)	(0.007)	(0.032)	(0.020)	(0.015)
IV-other	0.067	0.057	0.088**	0.064	0.084	0.170	0.037	0.063
	(0.063)	(0.053)	(0.038)	(0.064)	(0.064)	(0.115)	(0.064)	(0.071)
DID-IV	0.289***	0.290***	0.302***	0.273***	0.255***	0.316***	0.284***	0.237**
	(0.061)	(0.060)	(0.078)	(0.081)	(0.080)	(0.094)	(0.057)	(0.088)
DID-classic	0.321***	0.236***	0.305***	0.291***	0.243***	0.324***	0.350***	0.107
	(0.070)	(0.055)	(0.092)	(0.073)	(0.084)	(0.090)	(0.052)	(0.091)
Income Control (omitted: Auten Carroll)								
none	-0.210***	-0.205***	-0.200***	-0.201***	-0.212***	-0.203***	-0.207***	-0.210***
	(0.027)	(0.030)	(0.027)	(0.026)	(0.025)	(0.029)	(0.029)	(0.026)
Gruber Saez Spline	-0.018***	-0.010	-0.020***	-0.015*	-0.008	-0.016***	-0.020***	-0.011
	(0.005)	(0.008)	(0.005)	(0.008)	(0.010)	(0.004)	(0.004)	(0.009)
Kopczuk-type	-0.016**	-0.006	-0.020***	-0.017*	-0.009	-0.018***	-0.019***	-0.008
	(0.007)	(0.007)	(0.007)	(0.009)	(0.006)	(0.007)	(0.007)	(0.006)
other	-0.034*	-0.003	-0.030	-0.055*	-0.074**	-0.048**	0.006	-0.050***
	(0.020)	(0.008)	(0.021)	(0.027)	(0.028)	(0.002)	(0.010)	(0.017)
Difference Length (omitted: 3-years)								
1 year	0.069	0.033	0.079	0.066	0.067	0.069	0.042	0.065
	(0.074)	(0.051)	(0.098)	(0.074)	(0.068)	(0.094)	(0.054)	(0.068)
2 years	-0.010	-0.060***	-0.009	-0.017	0.002	-0.009*	-0.040	-0.009
	(0.025)	(0.016)	(0.026)	(0.019)	(0.010)	(0.025)	(0.027)	(0.007)
4 years and more	0.079	-0.018	0.081	0.066	0.044*	0.064	0.063	0.042
	(0.054)	(0.022)	(0.056)	(0.040)	(0.022)	(0.033)	(0.040)	(0.028)
Country Group (omitted: USA)								
Scandinavia		-0.128***						
		(0.027)						
Other Countries		0.082						
		(0.060)						
Additional Variables								
Intro top bracket			-0.032					
			(0.093)					
reduce brackets				-0.034*				
				(0.017)				
Gini Coefficient					0.010***			
					(0.003)			
Unemployment Rate						-0.009***		
						(0.006)		
Fraction of self-employed							0.018**	
							(0.008)	
modern taxes (in 2005)								-0.010***
								(0.002)
Constant	0.072***	0.189***	0.067***	0.096***	-0.167***	0.124***	-0.104	0.480***
	(0.006)	(0.029)	(0.007)	(0.017)	(0.058)	(0.098)	(0.074)	(0.098)
Country Group Dummy Var	no	yes	no	no	no	no	no	no
(Publication) Decade Dummy Var	no	no	no	no	no	no	no	no
Observations	852	852	836	836	852	776	846	852
Adjusted R^2	0.573	0.653	0.539	0.547	0.624	0.628	0.618	0.628

Columns (1) to (8) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For observations that are based on classic DID approach, I do not have information of the share of self employed that correspond to the respective mean year of observation.

Table 16: WLS before deductions - Contextual Variables

Dependent Variable: Income Elasticity BEFORE deductions	(9)	(10)
Reg. Technique (omitted: IV: Δ mechanical tax rate		
IV: lagged Δ mechanical tax rate	0.029***	0.038***
	(0.005)	(0.010)
IV-other	0.066	0.074
	(0.063)	(0.061)
DID-IV	0.266***	0.274***
	(0.079)	(0.075)
DID-classic	0.263***	0.289***
	(0.081)	(0.081)
Income Control (omitted: Auten Carroll)		
none	-0.213***	-0.213***
	(0.024)	(0.024)
Gruber Saez Spline	-0.010	-0.012
	(0.009)	(0.007)
Kopczuk-type	-0.011*	-0.013**
	(0.006)	(0.006)
other	-0.076**	-0.075*
	(0.032)	(0.038)
Difference Length (omitted: 3-years)		
1 year	0.061	0.067
	(0.066)	(0.071)
2 years	0.011	0.015
	(0.019)	(0.027)
4 years and more	0.060	0.074
	(0.036)	(0.049)
Additional Variables		
Top 10%	0.832***	
	(0.280)	
Top 1%		0.794**
		(0.361)
Constant	-0.143**	0.024*
	(0.064)	(0.013)
Country Group Dummy Var	no	no
(Publication) Decade Dummy Var	no	no
Observations	846	846
Adjusted R^2	0.626	0.610

E.2 Contextual Factors - After Deductions (AD) - Full Results

Table 17: WLS after deductions - Contextual Factors

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reg. Technique (omitted: IV: Δ mechanical tax rate)								
IV: (lagged) Δ mechanical tax rate	0.447***	0.327***	0.489***	0.149	0.438***	0.3361	0.435***	0.498***
	(0.129)	(0.116)	(0.125)	(0.135)	(0.131)	(0.241)	(0.122)	(0.128)
IV-other	-0.358***	-0.122	-0.359***	-0.011	-0.350***	-0.133**	-0.352***	-0.453***
	(0.121)	(0.141)	(0.121)	(0.103)	(0.125)	(0.124)	(0.120)	(0.116)
DID-IV	-0.677***	-0.399*	-0.695***	-0.492***	-0.677***	-0.254***	-0.651***	-0.598**
	(0.178)	(0.228)	(0.172)	(0.084)	(0.174)	(0.278)	(0.215)	(0.240)
DID-classic	0.129	0.278**	0.125	0.310**	0.139	0.945***	0.139	0.052
	(0.114)	(0.123)	(0.117)	(0.120)	(0.109)	(0.118)	(0.118)	(0.120)
Income Control (omitted: Auten Carroll)								
none	0.128	0.048	0.122	-0.095	0.124	0.043	0.103	0.127
	(0.126)	(0.135)	(0.129)	(0.090)	(0.138)	(0.182)	(0.115)	(0.129)
Gruber Saez Spline	-0.033	-0.086	-0.035	-0.222**	-0.040	-0.110	-0.029	0.005
	(0.105)	(0.125)	(0.107)	(0.087)	(0.121)	(0.140)	(0.108)	(0.134)
Kopczuk-type	-0.380***	-0.240**	-0.401***	-0.241**	-0.372***	-0.372***	-0.373***	-0.436***
	(0.114)	(0.101)	(0.118)	(0.091)	(0.107)	(0.113)	(0.113)	(0.123)
other	-0.327*	-0.166	-0.329*	-0.185	-0.316*	-0.402**	-0.331*	-0.411**
	(0.180)	(0.157)	(0.1853)	(0.113)	(0.163)	(0.178)	(0.163)	(0.192)
Difference Length (omitted: 3-years)								
1 year	0.026	-0.004	0.028	0.044	0.023	0.015	0.050	0.042
	(0.133)	(0.155)	(0.133)	(0.114)	(0.130)	(0.130)	(0.105)	(0.117)
2 years	0.199**	0.047	0.245**	0.074	0.186	0.186***	0.198	0.276***
	(0.077)	(0.075)	(0.097)	(0.093)	(0.114)	(0.057)	(0.159)	(0.083)
4 years and more	0.258	0.230	0.285*	0.096	0.246	0.004	0.268	0.422**
	(0.171)	(0.149)	(0.159)	(0.207)	(0.173)	(0.157)	(0.188)	(0.174)
Country Group (omitted: USA)								
Scandinavia		0.018						
		(0.092)						
Other countries		0.258***						
		(0.092)						
Additional Variables								
Intro top bracket			-0.086					
			(0.116)					
reduce brackets				-0.410***				
				(0.087)				
Gini Coefficient					0.002			
					(0.012)			
Unemployment Rate						-0.049		
						(0.041)		
Fraction of self-employed							0.002	
							(0.034)	
modern taxes (in 2005)								0.013
								(0.009)
Constant	0.443***	0.281**	0.448***	0.682***	0.393	0.071	0.417	-0.028
	(0.113)	(0.110)	(0.116)	(0.074)	(0.290)	(0.356)	(0.356)	(0.335)
Country Group Dummy Var	no	yes	no	no	no	no	no	no
(Publication) Decade Dummy Var	no	no	no	no	no	no	no	no
Observations	596	596	587	587	596	537	591	596
Adjusted R^2	0.580	0.597	0.563	0.656	0.579	0.646	0.600	0.590

Columns (1) to (8) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For observations that are based on classic DID approach, I do not have information of the share of self employed that correspond to the respective mean year of observation.

Table 18: WLS after deductions - Contextual Factors

Dependent Variable: Income Elasticity AFTER deductions	(9)	(10)
Reg. Technique (omitted: IV: Δ mechanical tax rate)		
IV: (lagged) Δ mechanical tax rate	0.111	0.236*
	(0.172)	(0.136)
IV-other	-0.213	-0.347**
	(0.127)	(0.130)
DID-IV	-0.550***	-0.659***
	(0.084)	(0.112)
DID-classic	0.290**	0.211*
	(0.134)	(0.120)
Income Control (omitted: Auten Carroll)		
none	-0.034	0.051
	(0.084)	(0.102)
Gruber Saez Spline	-0.195***	-0.108
	(0.066)	(0.084)
Kopczuk-type	-0.180	-0.244*
	(0.135)	(0.127)
other	-0.117	-0.196
	(0.174)	(0.175)
Difference Length (omitted: 3-years)		
1 year	0.025	0.016
	(0.119)	(0.129)
2 years	0.019	0.098
	(0.118)	(0.081)
4 years and more	0.102	0.168
	(0.242)	(0.233)
Additional Variables		
Top 10%	3.301***	
	(1.199)	
Top 1%		3.645**
		(1.689)
Constant	-0.583	0.115
	(0.412)	(0.189)
Country Group Dummy Var	no	no
(Publication) Decade Dummy Var	no	no
Observations	591	591
Adjusted R^2	0.651	0.616

F Sensitivity Analysis and Robustness Checks

F.1 Sensitivity Analysis

In this section, I limit the number of estimates along various dimensions: (i) I drop studies that are released prior to 2002, (ii) I consider only published articles or (iii) only US studies and (iv) I only consider taxable income elasticities. Results are presented in Table 19 and they vary slightly compared to the baseline results when I consider only published articles and only US studies. For US studies, the constant for BD elasticities is larger and smaller for AD elasticities compared to the baseline results shown in Table 2 and 3 (column 2).[26] Moreover, the degree of influence of other factors changes. The use of (lagged) mechanical tax rate changes lead to an increase of 0.541 compared to an approach that relies on mechanical tax rate changes as an instrument. On the other hand DID and DID IV does not make a big difference compared to an approach using the standard mechanical tax rate changes instrument. The coefficient of DID-classic is very large but mainly driven by older studies (reported < 2002).

Table 19: Sensitivity Analysis: Different Sample Restrictions

Dependent Variable:	drop studies prior to 2002		(only) Published		(only) US studies		(only) Taxable
Income Elasticity ...	BD	AD	BD	AD	BD	AD	Income
Reg. Technique (omitted: IV:Δ mech. tax rate))							
IV: (lagged) Δ mech. tax rate	0.050*	0.454***	0.043	0.342***	0.563***	0.305*	0.454***
	(0.025)	(0.129)	(0.029)	(0.046)	(0.141)	(0.174)	(0.127)
IV-other	0.046	-0.356***	0.074	-0.095	0.036	0.334***	-0.354***
	(0.062)	(0.123)	(0.045)	(0.165)	(0.118)	(0.124)	(0.121)
DID-IV	0.286***	-0.398***	0.323***	-0.408**	0.102	0.176**	-0.786***
	(0.063)	(0.123)	(0.047)	(0.146)	(0.133)	(0.078)	(0.209)
DID-classic	0.321***	0.091	0.375***	0.250**	0.087	1.521***	0.881***
	(0.070)	(0.115)	(0.043)	(0.092)	(0.102)	(0.154)	(0.138)
Income Control (omitted: Auten Carroll)							
none	-0.208***	0.116	-0.215***	-0.112	-0.174***	-0.249	0.103
	(0.028)	(0.129)	(0.026)	(0.144)	(0.055)	(0.172)	(0.170)
Gruber Saez Spline	-0.016***	-0.046	-0.015**	-0.078	-0.060	-0.015	-0.058
	(0.004)	(0.107)	(0.006)	(0.069)	(0.041)	(0.057)	(0.151)
Kopczuk-type	-0.015**	-0.392***	-0.012*	-0.273***	-0.170**	-0.116**	-0.405**
	(0.006)	(0.116)	(0.006)	(0.055)	(0.069)	(0.056)	(0.160)
other	-0.032	-0.334*	-0.016	-0.268**	-0.123	-0.073	-0.346
	(0.019)	(0.182)	(0.010)	(0.106)	(0.077)	(0.146)	(0.217)
Difference Length (omitted: 3-years)							
1 year	0.069	0.023	0.024	0.103	0.136	-0.113	0.023
	(0.074)	(0.135)	(0.047)	(0.129)	(0.101)	(0.072)	(0.131)
2 years	-0.010	0.181**	-0.038***	0.268***	-0.090	0.052	0.181*
	(0.025)	(0.083)	(0.012)	(0.056)	(0.150)	(0.137)	(0.095)
4 years and more	0.077	0.136	0.013	0.253	0.210**	0.037	0.164
	(0.055)	(0.196)	(0.013)	(0.211)	(0.094)	(0.130)	(0.161)
Constant	0.070***	0.456***	0.067***	0.330***	0.202***	0.280***	0.468***
	(0.005)	(0.114)	(0.006)	(0.056)	(0.043)	(0.072)	(0.159)
Observations	770	560	631	430	467	401	553
Adjusted R^2	0.579	0.582	0.705	0.648	0.111	0.389	0.580

Note: BD refers to the before deductions subsample and AD to the after deductions subsample. All results are based on Weighted Least Squares (WLS) with the inverse of an estimate's variance as analytical weights. The baseline specification involves only controls for estimation technique (regression technique, income control and difference length). Standard errors (in parentheses) are clustered at the study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[26] The results within the US subsample but also the baseline results remain remarkably robust even when I exclude all estimates extracted from Weber (2014*b*).

F.2 Robustness Checks: Different Estimation Techniques

The upper (lower) part of the Table displays results based on the BD (AD) subsample. Column (1) display the baseline results obtained in column (2) of Tables 2 and 3. In Column (2), I present results based on a random effects meta-regression technique. The weights in the baseline WLS represent only the within study variance and neglect any possible between study variance. In contrast the estimation used here, it is equivalent to the baseline WLS with an additive between study component in the denominator of the weights. Stanley and Doucouliagos, 2017 show that WLS is superior to conventional random-effects meta-regression estimation. In case of publication bias, in particular, WLS always reveals a smaller bias than the random effects model. Moreover, random effects estimates are highly sensitive to the accuracy of the estimate of the between study variance.

For illustration, results based on a simple OLS are presented in column (4). Since we observe large heteroscedasticity among estimates, an OLS procedure is never appropriate in a meta analysis. To increase efficiency, a WLS procedure is always preferable.

Column (5) shows results that are based on WLS with weights that are based on the inverse of the share of observations per study in relation to the full sample. Given that my collected sample does not consist only of one estimate per study but of all available estimates a particular study provides, there's a risk that the baseline results are driven only by a small number of studies that offer a lot of estimates.

It seems reasonable to assume that extracted estimates themselves are influenced by their sample size. For instance, a dataset that almost covers the entire population might produce a different estimate and standard error compared to a dataset of a few hundred observations. In column (6) I weight each primary estimate with the sample size of the respective study. The difference between those results compared to a standard WLS with precision as a weight should be small, since the sampling error is to large extent determined by the respective sample size.

The BD subsample is based on 38 studies and the AD subsample on 37 studies. To check whether clustering in the meta-analysis produces misleading inferences, I apply a wild-cluster bootstrap procedure proposed by Cameron et al. (2008) for improved inference with only few cluster (see Column (3)).

Table 20: Robustness Checks: Different Estimation Techniques

Dependent Variable: Income Elasticity BEFORE deductions	(1) WLS	(2) META	(3) WILD	(4) OLS	(5) EQUAL	(6) NOBS
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.050*	0.022	0.050	0.242	0.436	0.017
	(0.025)	(0.068)	(0.074)	(0.325)	(0.382)	(0.060)
IV-other	0.067	-0.120**	0.067	-0.227*	-0.398**	-0.101
	(0.063)	(0.056)	(0.072)	(0.131)	(0.155)	(0.113)
DID-IV	0.289***	0.067	0.289***	0.293	-0.192	0.386***
	(0.061)	(0.047)	(0.000)	(0.237)	(0.170)	(0.119)
DID-classic	0.321***	-0.078	0.321***	-0.583	-0.435**	0.088
	(0.070)	(0.295)	(0.000)	(0.384)	(0.178)	(0.114)
Income Control (omitted: Auten Carroll)						
none	-0.210***	-0.184***	-0.210***	0.255	-0.097	-0.223***
	(0.027)	(0.035)	(0.068)	(0.325)	(0.166)	(0.016)
Gruber Saez Spline	-0.018***	-0.184***	-0.018***	-0.363**	-0.163	-0.030**
	(0.005)	(0.035)	(0.006)	(0.134)	(0.221)	(0.015)
Kopczuk-type	-0.016**	-0.229***	-0.016***	-0.257*	-0.360*	-0.016
	(0.007)	(0.033)	(0.005)	(0.129)	(0.186)	(0.014)
other	-0.034*	-0.265***	-0.034	-0.282**	-0.414***	-0.172***
	(0.020)	(0.040)	(0.036)	(0.107)	(0.115)	(0.050)
Difference Length (omitted: 3-years)						
1 year	0.069	0.226***	0.069	0.175	0.303**	0.304**
	(0.074)	(0.030)	(0.113)	(0.145)	(0.141)	(0.112)
2 years	-0.010	-0.021	-0.010	-0.111	-0.096	0.121***
	(0.024)	(0.041)	(0.032)	(0.138)	(0.186)	(0.044)
4 years and more	0.079	0.053	0.079	-0.002	0.074	0.146**
	(0.054)	(0.038)	(0.110)	(0.164)	(0.146)	(0.054)
Constant	0.072***	0.287***	0.072***	0.408***	0.485***	0.073***
	(0.006)	(0.027)	(0.000)	(0.130)	(0.139)	(0.012)
Observations	852	852	852	852	852	788
Adjusted R^2	0.573		0.573	0.018	0.061	0.219

Income Elasticity AFTER deductions	(1) WLS	(2) META	(3) WILD	(4) OLS	(5) EQUAL	(6) NOBS
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.447***	0.288***	0.447***	0.361***	0.256***	0.414***
	(0.129)	(0.072)	(0.000)	(0.119)	(0.122)	(0.079)
IV-other	-0.358***	0.012	-0.358	0.113	0.243	-0.260***
	(0.121)	(0.055)	(0.235)	(0.117)	(0.212)	(0.042)
DID-IV	-0.677***	-0.171**	-0.677	-0.121	-0.307**	-0.398***
	(0.178)	(0.086)	(0.476)	(0.161)	(0.129)	(0.088)
DID-classic	0.129	0.653***	0.129	0.588*	0.791*	-0.092
	(0.114)	(0.066)	(0.166)	(0.343)	(0.436)	(0.083)
Income Control (omitted: Auten Carroll)						
none	0.128	0.083*	0.128	0.053	-0.318	-0.088
	(0.126)	(0.049)	(0.166)	(0.168)	(0.234)	(0.074)
Gruber Saez Spline	-0.033	0.031	-0.033	-0.058	-0.355	-0.266***
	(0.105)	(0.059)	(0.074)	(0.108)	(0.232)	(0.057)
Kopczuk-type	-0.380***	-0.070	-0.380***	-0.050	-0.196	-0.560***
	(0.114)	(0.049)	(0.122)	(0.101)	(0.197)	(0.085)
other	-0.327*	0.003	-0.327	-0.228	-0.552**	-0.474***
	(0.180)	(0.124)	(0.253)	(0.161)	(0.237)	(0.125)
Difference Length (omitted: 3-years)						
1 year	0.026	0.035	0.026	-0.014	0.225	0.050
	(0.133)	(0.041)	(0.079)	(0.143)	(0.146)	(0.042)
2 years	0.199**	0.137*	0.199	0.088	0.322**	0.069
	(0.077)	(0.079)	(0.139)	(0.174)	(0.130)	(0.098)
4 years and more	0.258	0.110	0.258	0.143	0.276**	-0.158***
	(0.171)	(0.067)	(0.243)	(0.139)	(0.094)	(0.041)
Constant	0.443***	0.244***	0.443***	0.282***	0.446***	0.660***
	(0.113)	(0.039)	(0.000)	(0.093)	(0.139)	(0.084)
Observations	596	596	596	596	596	567
Adjusted R^2	0.580		0.580	0.115	0.186	0.544

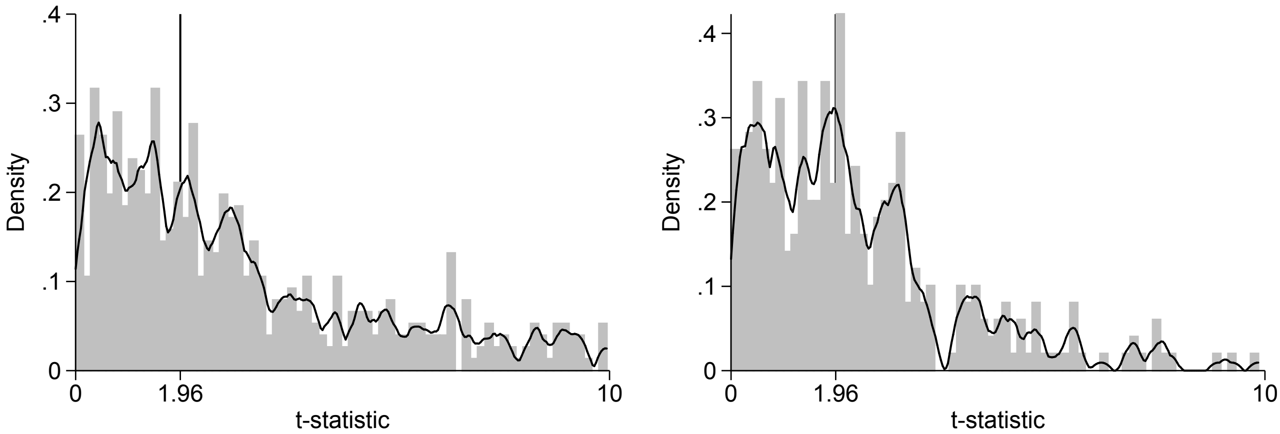
Except for column 3 standard errors (in parentheses) are clustered at the study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The sample size in column (6) is lower because the sample size is not observed for every primary estimate.

G Selective Reporting Bias: more information

G.1 Distribution of z-statistics - only with income controls

Figure 7: Distribution of z-statistics - only with income controls.



(a) Before Deductions (b) After Deductions

Note: The left (right) figure is based on the before (after) deductions subsample. The 5% significance value (=1.96) is highlighted.

G.2 Selective Reporting Bias: BD - Full Results

Table 21: WLS before deductions: Publication Bias Full Results

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)
Reg. Technique (omitted: IV: Δ mechanical tax rate)					
IV: lagged Δ mechanical tax rate	0.027*	0.026	0.025*	0.020	0.022
	(0.016)	(0.019)	(0.014)	(0.013)	(0.015)
IV-other	-0.167	-0.115	-0.170*	-0.222*	-0.200*
	(0.100)	(0.090)	(0.092)	(0.114)	(0.113)
DID-IV	0.191*	0.226**	0.208**	0.195*	0.196*
	(0.102)	(0.089)	(0.097)	(0.101)	(0.101)
DID-classic	-1.025***	-0.615**	-0.794**	-0.127	-0.144
	(0.305)	(0.287)	(0.306)	(0.217)	(0.343)
Income Control (omitted: Auten Carroll)					
none	-0.207***	-0.206***	-0.207***	-0.210***	-0.207***
	(0.029)	(0.030)	(0.029)	(0.027)	(0.029)
Gruber Saez Spline	-0.017***	-0.012*	-0.015***	-0.016***	-0.016***
	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)
Kopczuk-type	-0.016**	-0.009	-0.014**	-0.016**	-0.015***
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
other	-0.026	-0.009	-0.022*	-0.019	-0.024
	(0.016)	(0.009)	(0.012)	(0.012)	(0.016)
Difference Length (omitted: 3-years)					
1 year	0.044	0.034	0.039	0.032	0.039
	(0.061)	(0.053)	(0.059)	(0.054)	(0.058)
2 years	-0.021	-0.050***	0.016	-0.036***	-0.023
	(0.020)	(0.012)	(0.017)	(0.012)	(0.020)
4 years and more	0.055	-0.008	0.045	0.024	0.052
	(0.042)	(0.018)	(0.034)	(0.022)	(0.040)
Standard Error	3.725***	2.653***	4.203***	0.768	0.679
	(0.787)	(0.747)	(0.933)	(0.580)	(1.034)
Publish Type (omitted: published)					
Working Paper		0.411			
		(0.304)			
Std.Error* Working Paper		1.165			
		(1.249)			
Journal impact factor			-0.013		
			(0.009)		
Std.Error* Impact Factor			-0.051		
			(0.037)		
Dummy if obs > median(obs)				0.816***	
				(0.259)	
Std.Error*D if obs > median(obs)				4.467***	
				(1.128)	
Dummy reported prior to 2009					0.632**
					(0.309)
Std.Error*D reported prior to 2009					3.839***
					(1.454)
Constant	0.921***	0.670***	1.044***	0.449***	0.469**
	(0.180)	(0.171)	(0.213)	(0.127)	(0.198)
Observations	852	852	852	852	852
Adjusted R^2	0.618	0.656	0.631	0.645	0.630

Columns (1) to (5) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Included standard errors as explanatory variables are normalized. It allows an interpretation as standard deviation.

G.3 Selective Reporting Bias: AD - Full Results

Table 22: WLS after deductions Publication Bias Full Results

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)
Reg. Technique (omitted: IV: Δ mechanical tax rate)					
IV: lagged Δ mechanical tax rate	0.496***	0.450***	0.442***	0.511***	0.504***
	(0.128)	(0.118)	(0.090)	(0.132)	(0.136)
IV-other	-0.352***	-0.012	-0.199	-0.348***	-0.353***
	(0.118)	(0.130)	(0.125)	(0.118)	(0.112)
DID-IV	-0.581**	-0.518***	-0.467**	-0.768***	-0.630**
	(0.216)	(0.152)	(0.194)	(0.263)	(0.260)
DID-classic	0.144	0.277**	0.244**	0.149	0.019
	(0.110)	(0.107)	(0.097)	(0.123)	(0.153)
Income Control (omitted: Auten Carroll)					
none	0.117	-0.027	0.018	0.139	0.181
	(0.130)	(0.141)	(0.130)	(0.140)	(0.192)
Gruber Saez Spline	-0.031	-0.162	-0.117	0.010	0.031
	(0.109)	(0.120)	(0.112)	(0.132)	(0.166)
Kopczuk-type	-0.390***	-0.213**	-0.156*	-0.356**	-0.315
	(0.117)	(0.091)	(0.088)	(0.140)	(0.194)
other	-0.323*	-0.217	-0.042	-0.279	-0.257
	(0.183)	(0.148)	(0.194)	(0.190)	(0.231)
Difference Length (omitted: 3-years)					
1 year	0.028	-0.007	-0.003	0.042	0.045
	(0.132)	(0.154)	(0.152)	(0.118)	(0.105)
2 years	0.210***	0.242***	0.244**	0.234***	0.252***
	(0.073)	(0.041)	(0.113)	(0.064)	(0.071)
4 years and more	0.368**	0.094	0.321**	0.231	0.418**
	(0.175)	(0.156)	(0.141)	(0.168)	(0.164)
Standard Error	-0.226	0.276*	-0.926***	-0.494	-0.470
	(0.217)	(0.153)	(0.337)	(0.434)	(0.572)
Publish Type (omitted: published)					
Working Paper		-0.833*			
		(0.485)			
Std.Error*Working Paper		-1.398**			
		(0.589)			
Journal impact factor			0.038**		
			(0.015)		
Std.Error* Impact Factor			0.074***		
			(0.019)		
Dummy if obs > median(obs)				-0.258	
				(0.284)	
Std.Error*D if obs > median(obs)				0.074	
				(0.582)	
Dummy reported prior to 2009					-0.118
					(0.350)
Std.Error*D reported prior to 2009					0.128
					(0.681)
Constant	0.263	0.505***	-0.212	0.323	0.207
	(0.168)	(0.121)	(0.283)	(0.202)	(0.325)
Observations	596	596	596	596	596
Adjusted R^2	0.580	0.609	0.604	0.596	0.590

Columns (1) to (5) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Included standard errors as explanatory variables are normalized. It allows an interpretation as standard deviation.